

There's a lot of hype and misinformation about the new Google algorithm update. What actually is BERT, how does it work, and why does it matter to our work as SEOs? Join our own machine learning and natural language processing expert Britney Muller as she breaks down exactly what BERT is and what it means for the search industry. Hey, Moz fans. Welcome to another edition of Whiteboard Friday. Today we are talking about all things BERT and I'm super excited to attempt to really break this down for everyone. I don't claim to be a BERT expert but I've done lots and lots of research. --I've been able to interview some experts in the field and my goal is to try to be a catalyst for this information to be easier to understand.

There is a ton of commotion going on right now in the industry around BERT and how you can't optimize for it. While that is absolutely true, you cannot, you just need to be writing really good content for your users, I still think many of us got into this space because we are curious by nature. If you are curious to learn a little bit more about BERT and be able to explain it a little bit better to clients or have better conversations around the context of BERT, then I hope you enjoy this video. If not, and this isn't for you, that's fine too.

Word of caution: Don't over-hype BERT!

I'm so excited to jump right in. The first thing I do want to mention is I was able to sit down with Allyson Ettinger, who is a Natural Language Processing researcher. She is a professor at the University of Chicago and one of the kindest people. So appreciative that she took the time to chat with me about BERT.

My primary main takeaway from our lunch together was that it's very, very important to not over-hype BERT. There is a lot of commotion going on right now, but it's still far away from understanding language and context in the same way that we humans can understand it. So I think that's important to keep in mind that we are not overemphasizing what this model can do, but it's still really exciting and it's a pretty monumental moment in NLP and machine learning. Without further ado, let's jump right in.

Where did BERT come from?

I want to give everyone a wider context to where BERT came from and where it's going. I think a lot of times these announcements are kind of bombs dropped on the SEO industry, essentially, a still frame in a series of a movie but without the full before and after movie bits. We just get this one still frame. So we get this BERT announcement, but let's go back in time a little bit.

Natural language processing

Traditionally computers have had an impossible time understanding language. They can store text, we can enter text, but the understanding of language has always been incredibly difficult for computers. So along comes natural language processing (NLP), the field in which researchers develop unique models to solve for specific types of language understanding. A couple of examples are; named entity recognition, classification, sentiment analysis, and question and answering.

All of these have traditionally been solved by individual models fit to solve one specific language task and so it looks a little bit like your kitchen:

Think of the individual NLP models like utensils that you have in your kitchen, they all have one very specific task that they do very well.

Now consider a be-all-end-all kitchen utensil that is 11 of your most frequently used utensils in one. This is BERT, the one kitchen utensil that does eleven of the top natural language processing solutions really, really well after it's fine-tuned.

An exciting differentiation in the NLP space. That's why people are really excited about it because they no longer require all the individual models. --They can use BERT to solve for the majority of NLP tasks, which makes sense that Google would incorporate BERT into Google's algorithm.

Where is BERT going?

Where is this heading? Where is this going? Allyson had said,

"I think we'll be heading on the same trajectory for a while building bigger and better variants of BERT that are stronger in the ways that BERT is strong and probably with the same fundamental limitations."

There are already tons of different versions of BERT out there and we are going to continue to see more and more of that. It will be interesting to see where this space is heading.

How did BERT get so smart?

How about we take a look at a very oversimplified view of how BERT got so smart?

Google took Wikipedia text and a lot of money for computational power (TPUs in which they put together in a V3 pod) that can power these large models. They then used an unsupervised neural network to train from all of Wikipedia's text to better understand language and context.

What's interesting about how it learns is that it takes any arbitrary length of text (which is good because language is quite arbitrary in the way that we speak) and it transcribes it into a vector.

A vector is a fixed string of numbers. This helps language become translatable to a machine.

This happens in a really wild n-dimensional space that we can't even imagine. Putting similar contextual language into the same areas.

To get smarter and smarter BERT, similar to Word2vec, uses a tactic called masking.

Masking occurs when a random word within a sentence is hidden.

BERT being a bi-directional model looks to the words before and after the hidden word to help predict what the word is.

It does this over and over and over again until it's powerful in predicting masked words. It can then be further fine-tuned to do 11 of the most common natural language processing tasks. Really, really exciting and a fun time to be in this space.

What is BERT?

BERT is a pre-trained unsupervised natural language processing model. BERT can outperform 11 of the most common NLP tasks after fine-tuning, essentially becoming a rocket booster for Natural Language Processing and Understanding.

BERT is deeply bi-directional, meaning it looks at the words before and after entities and context pre-trained on Wikipedia to provide a richer understanding of language.

Check out this Whiteboard Friday for more context around what an unsupervised model is.

What are some things BERT cannot do?

Allyson Ettinger wrote this really great research paper called What BERT Can't Do. The most surprising takeaway from her research was this area of negation diagnostics, meaning that BERT isn't very good at understanding negation or what things are not.

For example, when inputted with a Robin is a... It predicted bird, which is right, that's great. But when entered a Robin is not a... It also predicted bird. So in cases where BERT hasn't seen negation examples or context, it will still have a hard time understanding that. There are a ton more really interesting takeaways in Allyson's research, highly suggest you check it out.

How do you optimize for BERT? (You can't!)

Finally, how do you optimize for BERT? Again, you can't. The only way to improve your website with this update is to write really great content for your users and fulfill the intent that they are seeking.

A great resource to help you understand and write better for NLP is Briggsby's On-page SEO for NLP article.

Google's growing capacity for Natural Question Understanding

One thing I just have to mention because I honestly cannot get this out of my head is this Keynote by Jeff Dean of Google. He's speaking about BERT and then goes into natural questions and natural question understanding. The big takeaway for me was this example

around, okay, let's say someone asked the question, "can you make and receive calls in airplane mode?"

Screenshot from Deep Learning for Solving Important Problems Keynote by Jeff Dean.

The block of text in which Google's natural language translation layer is trying to understand all this text is very technical and hard to understand:

Airplane mode, aeroplane mode, flight mode, offline mode, or standalone mode is a setting available on many smartphones, portable computers, and other electronic devices that, when activated, suspends radio-frequency signal transmission by the device, thereby disabling Bluetooth, telephony, and Wi-Fi. GPS may or may not be disabled, because it does not involve transmitting radio waves.

With these layers, and leveraging things like BERT, they were able to just answer "No" out of all of this very complex, long, confusing language. It's really, really powerful in our space.

Consider things like featured snippets; consider things like SERP features. I mean, this can start to have a huge impact in our space. So I think it's important to sort of have a pulse on where it's all heading and what's going on in this field.