What Is BERT? – Whiteboard Friday

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.

Click on the whiteboard image above to open a high-resolution version in a new tab!

Video Transcription

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. I have just 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 a little bit easier to understand.

There is a ton of commotion going on right now in the industry about you can’t optimize for BERT. 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. When I got to speak with her, the main takeaway 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 wanted 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 industry and it’s essentially a still frame in a series of a movie and we don’t get 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 understanding language has always been incredibly difficult for computers. So along comes natural language processing (NLP), the field in which researchers were developing specific models to solve for various types of language understanding. A couple of examples are named entity recognition, classification. We see sentiment, question

answering. All of these things have traditionally been sold by individual NLP models and so it looks a little bit like your kitchen.

If you think about the individual models like utensils that you use in your kitchen, they all have a very specific task that they do very well. But when along came BERT, it was sort of the be-all end-all of kitchen utensils. It was the one kitchen utensil that does ten-plus or eleven natural language processing solutions really, really well after it's fine tuned. This is a really exciting differentiation in the space. That's why people got really excited about it, because no longer do they have all these one-off things. They can use BERT to solve for all of this stuff, which makes sense in that Google would incorporate it into their algorithm. Super, super exciting.

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? I find this stuff fascinating. It is quite

amazing that Google was able to do this. Google took Wikipedia text and a lot of money for computational power TPUs in which they put together in a V3 pod, so huge computer system that can power these models. And they used an unsupervised **neural network.** What's interesting about how it learns and how it gets smarter is it takes any arbitrary length of text, which is good because language is quite arbitrary in the way that we speak, in the length of texts, and it transcribes it into a vector.

It will take a length of text and code it into a vector, which is a fixed string of numbers to help sort of translate it to the machine. This happens in a really wild and dimensional space that we can't even really imagine. But what it does is it puts context and different things within our language in the same areas together. Similar to Word2vec, it uses this trick called **masking**.

So it will take different sentences that it's training on and it will mask a word. It uses this bi-directional model to look at the words before and after it to predict what the masked word is. It does this over and over and over again until it's extremely powerful. And then it can further be fine-tuned to do all of these natural language processing tasks. Really, really exciting and a fun time to be in this space.

In a nutshell, BERT is the first deeply bi-directional. All that means is it's just looking at the words before and after entities and context, unsupervised language representation, pre-trained on Wikipedia. So it's this really beautiful pre-trained model that can be used in all sorts of ways.

What are some things BERT cannot do?

Allyson Ettinger wrote this really great research paper called What BERT Can't Do. There is a Bitly link that you can use to go directly to that. The most surprising takeaway from her research was this area of negation diagnostics, meaning that **BERT isn't very good at understanding negation**.

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. I highly suggest you check that out, really good stuff.

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.** And so you can't, but one thing I just have to mention because I honestly cannot get this out of my head, is there is a YouTube video where Jeff Dean, we will link to it, it's a keynote by Jeff Dean where he speaking about BERT and he 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? The block of text in which Google's natural language translation layer is trying to understand all this text. It's a ton of words. It's

kind of very technical, hard to understand.

With these layers, 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 just general **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.

I really hope you enjoyed this version of Whiteboard Friday. Please let me know if you have any questions or comments down below and I look forward to seeing you all again next time. Thanks so much.

Video transcription by Speechpad.com

http://tracking.feedpress.it/link/9375/12961338?utm_medium=feed&utm_source=feedpress.me&utm_campaign=Feed%3A+mozblog