How Search Engines Use Machine Learning: 9 Things We Know for Sure by @_kevinrowe

When we first started hearing about machine learning in the early 2010s, it seemed scary at first.

But once it was explained to us (and we realized how technology is already being used to provide us with solutions), we started to get down to the practical questions:

- How are search engines using machine learning?
- How will it affect SEO?

Machine learning is essentially using algorithms to calculate trends, value, or other characteristics of specific things based on historical data.

Google has even declared itself a machine learning-first company.

If you want to learn more about the tactical side of this technology, Eric Enge has a great write-up on Moz explaining how machine learning impacts SEO from a mathematical standpoint.

Search engines like to always experiment with how they can use this evolving technology, but here are nine ways we know that they are currently using machine learning and how it relates to SEO or digital marketing.

1. Pattern Detection

Search engines are using machine learning for pattern detections that help identify spam or duplicate content. They plugged in common attributes of low-quality content, such as:

- The presence of several outbound links to unrelated pages.
- Lots of uses of stop words or synonyms.
- Other such variables.

Being able to detect these kinds of patterns drastically cut down on the manpower it takes to review everything by actual people.

Even though there are still human quality raters, machine learning has helped Google automatically sift through pages to weed out low-quality pages without an actual human having to look at it first.

Machine learning is an ever-evolving technology, so the more pages that are analyzed, the more accurate it is (in theory).

2. Identifying New Signals

According to a 2016 podcast done with Gary Illyes from Google, RankBrain not only helps identify patterns in queries, it also helps the search engine identify possible new ranking signals.

These signals are sought after so Google can continue to improve the quality of search query results.

Illyes also mentioned in the podcast episode that more of Google's signals may become machine learning-based.

As search engines are able to teach technology how to run predictions and data on their own, there can be less manual labor and employees can move toward other things machines can't do, like innovation or human-centered projects.

3. It's Weighted as a Small Portion

However, even though machine learning is slowly transforming the way search engines find and rank websites, it doesn't mean it has a major, significant impact (currently) on our SERPs.

In the same podcast interview, Illyes says that it’s just part of their overall ranking signal platform, and is weighted as a small portion of their overall algorithm.

Google’s end goal is to use technology to provide users with a better experience. They don’t want to automate the entire process if that means the user won’t have the experience they are looking for.

So don’t assume machine learning will soon take over all search ranking; it is simply a small piece of the puzzle search engines have implemented to hopefully make our lives easier.

4. Custom Signals Based on Specific Query

Machine learning in search engines may vary depending on the query category or phrasing, according to a July 2017 study done at the University of Washington.

Researchers used Russian search engine Yandex to analyze results for different queries. They found that the types of results displayed depended largely on the query category or phrasing.

This means that machine learning can place more weights on variables more or less heavily in certain queries over others.

Overall, it was found that personalized searches customized by machine learning increased the click-through rate (CTR) of results about 10 percent.

As the user entered more queries into Yandex, it was found that the CTR continued to increase.

This is likely because the search engine was “learning” about that specific user’s preferences and could base its

information on past queries to present the most interesting information possible.

An example of this that is often used in conference presentations is a string of queries in one sitting and how the results change depending on what you last searched.

For instance, if I search “New York Football stadium” in an incognito browser, I get the answer of “MetLife Stadium.

Next, if I search in the same browser for just “jets,” Google is assuming that because my last query was about a football stadium, then this query is also about football.





As I continue my search, Google learns when I’ve turned into something else.

Searching for “Jaguars” in the same browser will bring up information about the NFL team the Jacksonville Jaguars (related to my last two searches).

But the instance I search “Zoo near San Diego” then start to type “zoo” again in the query box, Google suggests “zoos with jaguars” even though I haven’t searched jaguars a second time.



Search history is just one component of the search experience that machine learning uses to provide better results.

5. Image Search to Understand Photos

Back in 2013, it was reported that Flickr users upload 1.4 million photos per day, 40 million are uploaded to Instagram,

and Facebook users were uploading 350 million.

While these statistics have likely gone up (it was difficult to find more recent data), it shows that volume of photos that need to be cataloged and analyzed on the web daily.

This task is perfect for machine learning because it can analyze color and shape patterns and pair that with any existing schema data about the photograph to help the search engine understand what an image actually is.

This is how Google is able to not only catalog images for Google Image search results, but also powers its feature that allows users search by a photo file (instead of a text query).

Users can then find other instances of the photo online, as well as similar photographs that have the same subjects or color palette and information about the subjects in the photo, as in this example of a classic Christmas movie still:



The way the user interacts with these results can shape their SERPs in the future.

6. Identifying Similarities Between Words in a Search Query

Not only does query data get used by machine learning to identify and personalize a user's later queries, it also helps create patterns in data that shapes the search results other users are getting.

Google Trends is a great front-facing example of this. A phrase or word that doesn't mean anything initially (e.g. "planking" or "it's lit") may have nonsensical search results.

However, as its phrasing (and therefore, user searches) is used more over time, machine learning is able to display more

accurate information for those queries.

As language develops and transforms, machines are better able to predict our meanings behind the words we say and provide us with better information.

7. Improve Ad Quality & Targeting for Users

According to Google U.S. patent US20070156887 and US9773256 on ad quality, machine learning can be used to improve an "otherwise weak statistical model."

This means that Ad Rank can be influenced by a machine learning system.

"Bid amount, your auction-time ad quality (including expected clickthrough rate, ad relevance, and landing page experience), the Ad Rank thresholds, the context of the person's search" gets fed into the system on a keyword-by-keyword basis, to determine what thresholds are considered by Google for each keyword.

8. Synonyms Identification

When you see search results that don't include the keyword in the snippet it's likely due to Google using RankBrain to identify synonyms.

When searching for [phd degree] you'll see various results with the word "doctor" or "doctoral" as they can be used, for many degrees, interchangeable.

Google even highlights the synonyms in some cases, this time with "phd degrees," further indicating that it's recognizing the synonyms.



9. Query Clarification

One of my favorite subjects is search query user intent.

Users may be searching to buy (transactional), research (informational), or find resources (navigational) for any given search. Furthermore, a keyword could be useful to one or any of these intents.

By analyzing click patterns and the content type that users engage with (e.g. CTRs by content type) a search engine can leverage machine learning to determine the intent.

An example can be seen with the query “best college” in a Google search. The results are reviews and list of colleges all in one SERP, with the universities listed at the top.



Summary

While machine learning isn’t (and probably never will be) perfect, the more humans interact with it, the more accurate and “smarter” it will get.

This could be alarming to some – bringing visions of Skynet from the “Terminator” movies – however, the actual result is likely a better experience with technology that gives us the information and services we need, when we need it.

More Machine Learning Resources:

Image Credits

All screenshots taken by author, February 2018.

http://tracking.feedpress.it/link/13962/8364928