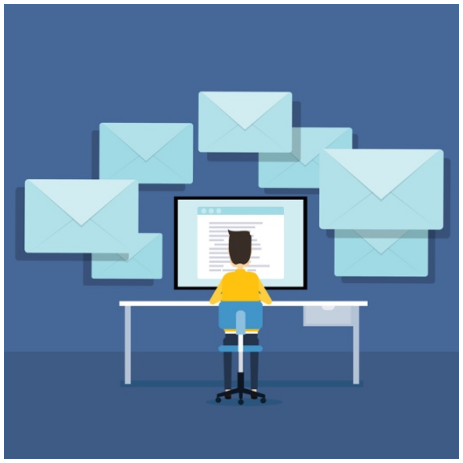
cortical.io

THE IMPACT OF EMAIL CLASSIFICATION ON YOUR BOTTOM LINE

How Hardware-Accelerated Natural Language Understanding Can Help

Vienna, November 2019



Email management is becoming more challenging every year. In 2019, business email accounts for more than 128.8 billion emails sent and received per day, according to the Radicati Group.[1] Adding to the challenge, many emails never make it to the right business account because they were sent to bulk accounts like info@company.com or sales@company.com.

Regardless of where the emails land, the average full-time worker spends 28 percent of the workday reading and answering email, according to a McKinsey analysis.[2] That amounts to a staggering 2.6 hours spent each day.

Corporate email continues to rule in the business world, but the deluge is impairing productivity, not to mention becoming unmanageable from a corporate perspective. TechJury[3] cites that at least five out of 10 emails are spam in 2019, but spam isn't the only problem. Oftentimes, the right email doesn't reach the right person within the organization quickly, resulting in operational delays and rising costs.

TOO LITTLE BUSINESS VALUE EXTRACTED FROM EMAILS

Despite these drawbacks, emails remain an extremely rich and essential source of business information. However, extracting insights and relevant information from emails is difficult and time-consuming. As the amount of emails continues to grow, it becomes even more difficult, time-intensive, and expensive to realize the business value.

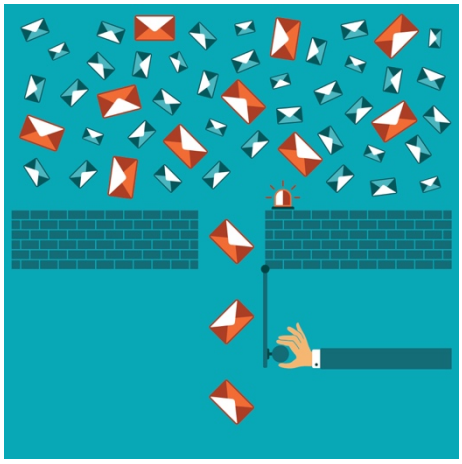
The main issue is that emails fall into the category of unstructured data that is typically text heavy. This results in irregularities and ambiguities that make it difficult to automate the process of analyzing and categorizing using traditional programs.

[1] https://www.radicati.com/wp/wp-content/uploads/2015/02/Email-Statistics-Report-2015-2019-Executive-Summary.pdf

[2] https://hbr.org/2019/01/how-to-spend-way-less-time-on-email-every-day

[3] https://techjury.net/stats-about/how-many-emails-are-sent-per-day/

MOST CURRENT EMAIL CLASSIFICATON SYSTEMS FALL SHORT



Most current technologies for email classification fall into two broad categories. Rule-based approaches classify text into organized groups by using a set of handcrafted linguistic rules. These rules instruct the system to use semantically relevant elements of a piece of text to identify relevant categories based on its content. Essentially, the system counts the number of word appearances in the text and subsequently classifies the content into predefined categories.

This approach has several disadvantages. For starters, these systems require deep knowledge of the domain. They are also time-consuming, because generating rules for a complex system can be quite challenging and usually requires a lot of analysis and testing. Rule-based systems are also difficult to maintain and don't scale well given that adding new rules can affect the results of the pre-existing rules.

Instead of relying on manually crafted rules, text classification systems based on Natural Language Processing (NLP) use statistical modeling. These approaches are usually more accurate than rule-based approaches; however, they involve building large-scale statistical models that are trained through machine learning (ML) using massive amounts of training data. In many instances, sufficient amounts of training data are not readily available. The process is slow, costly, and complex, involving several iterations of data preparation, feature engineering, model training, hyperparameter tuning, and evaluation. Even if more data is thrown at the model so it can become more accurate, there is still about 20 percent of text-based information that the model cannot precisely categorize.

Additionally, most standard ML-based systems cannot explain why they made a certain decision. This "black box" effect is a substantial drawback in light of regulations such as the EU's General Data Protection Regulation (GDPR), which stipulates that consumers have a right to know why and how corporations made decisions affecting them. Another issue with "black-box" models is that they are hard to improve. If their predictions are wrong but there is no way to explain why, it is difficult to know how to fix the model so that it delivers the correct output.

SEMANTIC FOLDING: A POWERFUL NEW APPROACH

Now, there is an out-of-the-box solution for mining emails for business content and classifying them, that comes as a pre-trained appliance with standard filters. If customized filtering is needed, it can be done with a fraction of the training data, as well as much higher accuracy and efficiency compared to other models. It is based on Semantic Folding, an innovative approach to Natural Language Understanding (NLU). Semantic Folding is a procedure for encoding the semantics of natural language text in a sparse distributed representation called a semantic fingerprint. This unique approach provides a framework for analyzing unstructured data such as emails similar to how the human brain processes language data.

Semantic Folding is inspired by and based on the principles of cerebral processing and neuroscience. By applying neuroscience tenets discovered by scientist and author Jeff Hawkins to the realm of NLU, Cortical.io can identify and analyze language faster and more accurately.

Using Semantic Folding, an artificial intelligence (AI)-driven algorithm can quickly categorize emails. The result is an intelligent, efficient way to analyze emails so organizations can tap the value they contain, even across multiple languages. More information on Semantic Folding can be found here.

MULTIPLE BUSINESS BENEFITS

Organizations that can figure out a way to efficiently categorize and segment emails can improve their bottom lines, reduce costs, improve productivity, streamline and automate business processes, and serve customers better. The potential results are tremendous, as well as applicable to any business function. For instance, organizations can:

- **Detect and surface only relevant emails** having to do with RFPs, bids, budgets, forecasts, and reporting for the Finance Department.
- **Assist sales and marketing by extracting emails** regarding market research, advertising, and customer satisfaction.
- **Surface emails about scheduling production**, ensuring product quality, and minimizing production costs to streamline manufacturing.

A CASE STUDY IN CUSTOMER SERVICE

Customer service is among the most compelling use cases for email classification. Cortical.io, in partnership with PwC Germany, worked with a large international shipping and container transportation company whose goal was to improve customer service by reducing the wasted efforts of handling irrelevant mails. The company sought to accelerate resolution of customer service requests, in 35 different languages, by keeping all no-case mails away from the case management system—without losing any case-relevant mail.

PwC applied Cortical.io's AI-based Semantic Folding approach to create a scalable, secure web service that detects and classifies emails with a high degree of accuracy. Only 50 percent of the emails that made it through the spam filters were relevant to a customer service case, causing wasted time and effort. Many of the non-relevant emails were automated confirmation messages providing updates about the status of shipments, such as a container left at a certain port, or a container that was not loaded onto the intended ship.

With the new solution, which was deployed in just eight weeks, customer service representatives were able to ignore irrelevant emails and focus on solving requests more efficiently based on case-relevant communications. The solution required very little training data, resulting in tremendous cost savings. It has the potential to save the company € 4.9 million annually.

TAMING THE EMAIL TORRENT WITH HARDWARE ACCELERATION

The solution developed for the transportation company employed Cortical.io's software, but it ran on standard hardware. Now imagine combining Cortical.io's software with hardware acceleration to process enormous streams of emails with orders of magnitude faster performance.

It is possible now, because Cortical.io offers the Messaging Classification Appliance. Delivered in partnership with Xilinx, Inc. and Super Micro, the Messaging Classification Appliance is a pre-packaged software and hardware solution that can filter, classify, and route streams of messages in real time at a

massive scale by understanding the semantic content – the meaning and intent of the messages.

The Messaging Classification Appliance is Cortical.io's first solution in the category of Semantic Supercomputing, a powerful approach that combines Semantic Folding with hardware acceleration to process and analyze natural language.[4] The appliance is powered by Xilinx FPGA-based Alveo Accelerator cards, which are well suited to Cortical.io's Semantic Folding algorithms and can speed up operations dramatically. Accurate results are delivered quickly—without the need for humans to manually sift through emails to determine what is pertinent. And, the solution works with and across any language, with little additional effort. The new solution is delivered with pre-configured hardware preloaded with Cortical.io software that can be integrated with an enterprise's IT infrastructure through REST APIs.

CONCLUSION

Although email is a faster, better alternative to paper documents and faxing, it can have a direct impact on corporate bottom lines by distracting workers from role-relevant tasks to spend hours daily dealing with unimportant messages. Considering this drag on workplace efficiency, it is time to adopt solutions for categorizing massive volumes of email in near real time using a powerful, low-cost, accurate email classification solution. The good news is that this type of solution can also be applied to a broad range of messages, including instant messages, tweets, and blog posts. The results are well worth it, including lower costs, higher productivity, improved customer satisfaction, and increased efficiency and responsiveness for all functional areas of the organization.

For more information about Messaging Classification and Cortical.io, visit www.cortical.io or contact info@cortical.io.

[4] https://www.cortical.io/static/downloads/semantic-supercomputing-white-paper-2019.pdf