A NEW FRONTIER FOR AI: THE FUTURE OF DEEP LEARNING

BY KJS COSTELLO

Industry thirst for artificial intelligence (AI) has been well documented over the last decade. While AI has been readily pursued since the 1960s, actual and perceived limitations caused interest to languish through the 70s and 80s, and enthusiasm was relegated to the world of science fiction. Pessimism remained rife until around 2008, since abated, however, by advances in computational power, increased tech investment, and numerous real-world accomplishments.

The potential of machine learning is once more conspicuous, fueling developer and investor enthusiasm alike. The growing real-world applications are out there for all to see – from self-driving cars to the advancement of speech recognition software – and with the rise in computational power, limitations are quickly dissipating. What was once unimaginable is quickly becoming possible, and that which we once confined to science fiction is becoming not only feasible, but tangible.

Deep learning, however, as a subset of machine learning (in turn a subset of AI) does not necessarily seek to establish an artificial intelligence, at least not AI as it is popularly understood. But it is AI as an unfettered intelligence, capable of learning independent of supervised input. Deep learning sets out to explore what it means to think, facilitating an unprecedented degree of autonomy within how a computer responds to new information data. For Persistent, these trends have exciting implications within the area of text analysis, inspiring a system able to understand the principles of syntax, semantics, and phonetics as it applies to language – for a system to view language as more than a collection of symbols, words, and sentences, and ultimately to derive meaning.

Over the past six years, industry enthusiasm for deep learning has been framed by the rise of Big Data. Beyond its buzzword status, Big Data is an incredibly valuable resource, which increasingly exists in abundance, especially when paired with intelligent learning algorithms. Superseding a human's capacity to interpret Big Data, deep learning has huge implications for the advancement of pattern detection, and in understanding the nitty-gritty details and connections between individual units of data. Naturally, for companies and markets, the potential is huge, whether in identifying consumer trends or predicting sales figures, etc. After all, data by itself achieves nothing.

If we consider all learning to be the result of human experiences, from which we derive our lessons and thought patterns, then Big Data represents a collation of innumerable experience. The hustle and bustle around Big Data has long awaited advancement in machine learning, to effectively analyze and derive patterns from the information that is gathered – even to learn.

So what is deep learning?

Deep learning depends upon deep neural networks, inspired by the activity within the human neocortex. Previously, machine learning could be better understood as supervised learning, which relies upon statistical or mathematical techniques to construct a model or system. Deep learning, however, explores the relationship between data to calculate analogous outcomes via a multi-layered neural network. As opposed to traditional machine learning methods, deep learning is able to detect subtle nuisance within the data presented, such as the spatial relationship of objects within a picture, and extrapolate details to discover relationships between similar data samples. Put simply, the software learns to recognize complex, deep-level patterns in digital representations of sounds, images, and words, as well as other data.

In areas such as image recognition, this has resulted in drastically reduced error rates. Between 2011 and 2015, demonstrated error rates across various technologies utilizing deep learning fell from a best-possible 25.8% to a best-possible 5%.

Deep learning has far-reaching implications within a range of industries, but the technology is specifically being pursued within target technologies, such as self-driving cars. It presents an otherwise unprecedented level of sophistication within image recognition, able to overcome the sequential decision variables (or sub-trees) derived from the mounting data that occurs during a car journey. The capacity in deep learning to handle an automated and continuous 'stream' of data presents a scalable and cross-industry solution to Big Data analytics, to match the fluidity and dynamism of modern digital experiences (such as, for example, tracing a user's consumer interest to match their current and ongoing need, versus a need that existed a week prior).

The ability to match a human's stream of consciousness had eluded the capacity of AI until recently. Deep learning's ability to continuously derive patterns from data streams, therefore, similarly creates new possibilities within speech and text analytics. Significantly, it can apply semantic meaning to combinations of words and sentences, to the ultimate end of mapping variable languages.

How does it work?

Research into neural networks began as early as the 1970s, but the network's ability to only process limited neurons, or layers, simultaneously prevented detection of complex patterns, and ultimately produced poorer results than alternative machine learning methods being explored at the time.

A key inhibitor was computational power. Multiple layers, or layered subroutines, are necessary to detect complex and nuisance patterns. Past methods of machine learning can be regarded as 'shallow' in so far that the patterns identified are surface layered (i.e. the algorithm has only been passed through one or two layers), resulting in inaccurate outcomes within the likes of image recognition and speech detection where there are unaccountable variations within different images, such as different shapes of the same animal.

Obviously, real neural networks are hopelessly more complex than we could hope to emulate, involving multimillion dimensions. However, these artificial networks are based on the abstract notion of a neuron; the basic principle of connected neurons exchanging messages in a fluid stream has inspired a significant part of deep learning. In broad strokes, these neurons have numerical 'weights' applied, which can be tuned based on experience, and which in turn facilitate the 'learning' aspect of neural networks by making the neural nets adaptive and dynamic. Weights determine how each neuron responds to a digitalized feature, such as an edge, a shade of color, or a symbol.

For example, if we had a collection of images of different household rooms, such as a bedroom, bathroom, and kitchen, deep learning could automatically infer that the bedroom, without necessary knowing it's called a bedroom, would have a bed, lamp, and wardrobe. It calculates this by figuring out a hierarchy of features that should be present in order for it to be included into the bedroom category, as well as the spatial relationship between the objects present. The layers in these scenarios, additional to the first input layer, could correspond, for example, to the shape of furniture, their distance apart and position, or the size of the room, etc. This pattern identification is entirely autonomous, and the categories or 'topics' of rooms will be calculated regardless of whether the system knows it is identifying a bedroom or a kitchen. In cases of inaccurate outcomes, the weights applied to the neurons can be modified in a process known as backpropagation, whereby small steps backwards are taken to increase the likelihood of a correct outcome.

Deep learning machines compile a web of weighted correlations, through which words and phrases are associated with their likely counterparts.

Persistent's role within deep learning

Persistent has invested heavily in deep learning, both in a contributory capacity and to position themselves as an industry leader in advanced text analytics. Machine learning is necessarily pervasive, driving the direction of many companies' research and innovation strategies.

But the exploration of deep learning is, fundamentally, an endeavor to problem solve. To this end, Persistent has specifically targeted deep learning within text analytics, as a means of revolutionizing (or, at least, reassessing) how machine learning improves the response to different languages. Importantly, this technology seeks to assuage the problems afforded by multiple and eclectic languages by removing the language as the dominant factor. Through automatic topic discovery, a system can respond dynamically to different languages, by identifying and 'learning' keywords and topics. To this end, previous rule-based systems have been inflexible and, invariably, have fallen behind. A system independent of the language, capable of adapting and learning, is necessary to mitigate the resource burden of different mappings for different languages. To this end, Persistent is first targeting the 24 Indian languages, to firstly highlight the problems presented by cross-development over multiple regional languages, and secondly to prove that the quantity of languages can, in fact, cease to be an obstacle within a deep learning system.

India's integration within digital society is inhibited by both the quantity of languages and the likelihood of individuals only ever acquiring one Indian language. More often than not, the type of language spoken corresponds to an individual's region of origin. This creates an intrinsic problem for traditional search functions, which are restricted to static like-to-like comparisons of character and symbol combinations, without being able to incorporate the meaning of the word, or even synonymous, into the search parameters. In general, search functions, as with traditional machine learning, text crawling, text analytics, and even speech recognition relies heavily on rule-based systems that make linear, surface-layer associations without the ability to divine semantics, apply meaning, or incorporate context into the search criteria.

To meet our design objectives, we first concentrated specifically on Hindi. Our first-layer input took the form of hard-pressed text, in this case texts from the Utter Pradesh (UP) Legislative Assembly. The text data was obtained through obstacle character recognition (OCR), which creates a digitalized, numerical representation of text, in this case Hindi characters. In this way, the nature of the data is irrelevant as all deep learning systems calculate data from digital representations. The computer does not know or recognize, for example, the Hindi language as anything other than a sequence of syllables; it could, in practice, be any other language.

The deep learning system uses automatic topic discovery to tag each sentence within the legislative documents to a specific topic. Topics are identified by the sequence of syllables, which, for example, could relate to terrorism or public health. The system will not necessarily know what these topics denote, per se, but will still be able to categorize the information accordingly. Secondly, the system will be able to derive the main keywords for each topic from the information gathered, and extract the summary of the topic. In cases of newly input sentences, which haven't been entered before, the system (through analogy and main keyword frequency) will allocate it to the closest topic. This step is essential because of its ability to derive meaning from sentences, and allocate them directly (through literal like-to-like comparison) and indirectly (through keyword density, topic, and ultimately meaning) to similar sentences. In this way, language ceases to be an abstract combination of characters and syllables; multilayered deep learning is able to penetrate language to its nuisance structure, in order to identify instances in which, for example, a word is used in the stead of another similar word (as a synonym) by identifying the words at either side. Without this, we are left with rudimentary keyword search functions, which fail to extract meaningful information and are unable to account for context.

Deep learning, however, differs significantly from summarization technologies. This is much more than counting instances in which a certain keyword appears; Persistent's deep learning system seeks to retain the meaning and structure of a sentence precursory to identifying it with a similar sentence, as opposed to viewing it as a 'bag-of-words'. Traditionally, bag-of-word summarization approaches could only account for exact instances of a word or sentence recurring, and could not account for misspelt instances or out-of-context uses; sentences would simply be flagged if they contained a certain density of keywords. Naturally, this resulted in extracted sentences having lost their semantic meaning; summarization technology does not care for fluidity or structure, only frequency. It assembles jigsaw pieces, but omits the overall picture.

Automatic topic discovery is facilitated by a quasi-deep learning bag-of-words method called LDA (Latent Dirichlet Allocation). The term bag-of-words is given to denote the literal process involved whereby words and sentences are assembled via their topic to derive a base similarity between like-to-like samples, but which ultimately lose syntax and fluidity within the end result (presented in a jumbled-up fashion). Still, it remains an important and effective method in topic discovery and allocation, and contributes to the bedrock of numerous machine learning technologies. In simple terms, it can be considered a first-step towards truer, more complicated deep learning.

But back to the problem. While topic discovery is an important aspect of text detection, deep learning goes beyond the LDA, bag-of-words approach (which, again, represents documents as mixtures of topics that produce words with certain probabilities). We also care about structure and sequence, using the neural network in combination with a sliding-window detection method. It does not only focus on top words (verbs and nouns), as previous methods did, but the connecting words that contribute to the sentence's meaning and fluidity (in this case, the conjugation of the infinitive 'to be'). If we search 'waffles are great', the only first-layer input can be 'are', through which the system would be able to predict the other two words. Likewise, 'waffles' and 'great' will provide the middle word, which is 'are'. This is calculated by applying a numeric vector, or weight, for each word that exists in a corpus, through which analogous associations are calculated, identifying the recurring relationship between particular units of data. As with the images, the numeric vector is a digitalized representation of symbols, sounds, images, etc.

The ultimate goal, therefore, is meaning extraction, and incorporating the meaning of a word or sentence into the search parameters. With a small amount of preprocessing, the system represents an adaptable backbone learning system that can accommodate any chosen language. Persistent has already demonstrated the system with Hindi, as well as four other Indian languages through the same sample data.

Just as deep learning is being pursued within cancer research, to tackle the continuously-evolving nature of cancers, so too can it be used to automatically adapt to changes in languages, or new languages themselves. The dynamism of the system is an essential part of its offering; preprocessing ends with the input language, through which the system can perform the exact same processes to reach similarly unprecedented accuracy in text analytics with only a few additional tweaks. Neither the quantity of symbols, the complexity of the language, nor the quantity of languages selected will inhibit the system's performance. This is an industry-first achievement in the pursuit of deep learning within text analytics, and will enable us to aggressively pursue markets and applications within India, before extending its application elsewhere.