De-Demonizing AI Regarding Localization

Andrzej Zydroń
CTO, XTM International
Gerrards Cross
azydron@xtm.cloud

Abstract

AI gets a lot of attention generally due to the stunning results that can be achieved, in fields such as medicine, automotive technology, diagnostic systems and of course translation. AI systems can seemingly outperform human beings in a wide range of tasks, from playing Chess, Go or even poker, to face and voice recognition. What is often lacking though, is a more realistic understanding of what intelligence is and the actual limitations that exist given the computing tools at our disposal.

The reality is much more prosaic: most of the mathematical basis of what is termed AI is not complicated and generally rooted in early 18th century mathematics, namely work done by Euler and Bayes.

Although some of the achievements of AI based systems may seem phenomenal, they are achieved through processing of gigantic amounts of data which would normally be beyond human capability.

How can we define intelligence? The definitive explanation of intelligence, specifically with regard to AI, was provided in 2005 by Jeff Hawkins in his seminal work: On Intelligence[1]. Intelligence is defined as comprising three key elements:

1. The ability to recognize patterns
2. The ability to store the memory of those patterns
3. The ability to predict future events based on the stored patterns

Using Hawkins' definition concerning AI we can classify AI systems according to the following criteria:

1. Expert Systems[2]
   a. Knowledge based[3]
   b. Rule based[4]
2. Machine Learning (ML)[5]
   . Bayesian based ML[6]
   Neural Networks based ML[7]

Expert systems rely on analysing the type of approach an expert in a given system uses to solve a given problem. We note the patterns and write a program that uses these patterns to solve a given problem. For example, to provide an optimal layout for a table in a browser page based on HTML markup we can use the same rules that a typographer would use to provide an elegant table layout. The rules are based on an optimal proportion of column headings to average column content's width bearing in mind the number of columns etc.

Knowledge Based systems rely on building a table of 'if..then..else..' rules that are typical in diagnostic systems, be it motor vehicle mechanics or medical diagnosis. For example: if the engine will not start then first check the amount of fuel, then the fuel pump function, battery level etc. Here the rules are in the form of a table rather than as part of the actual program.

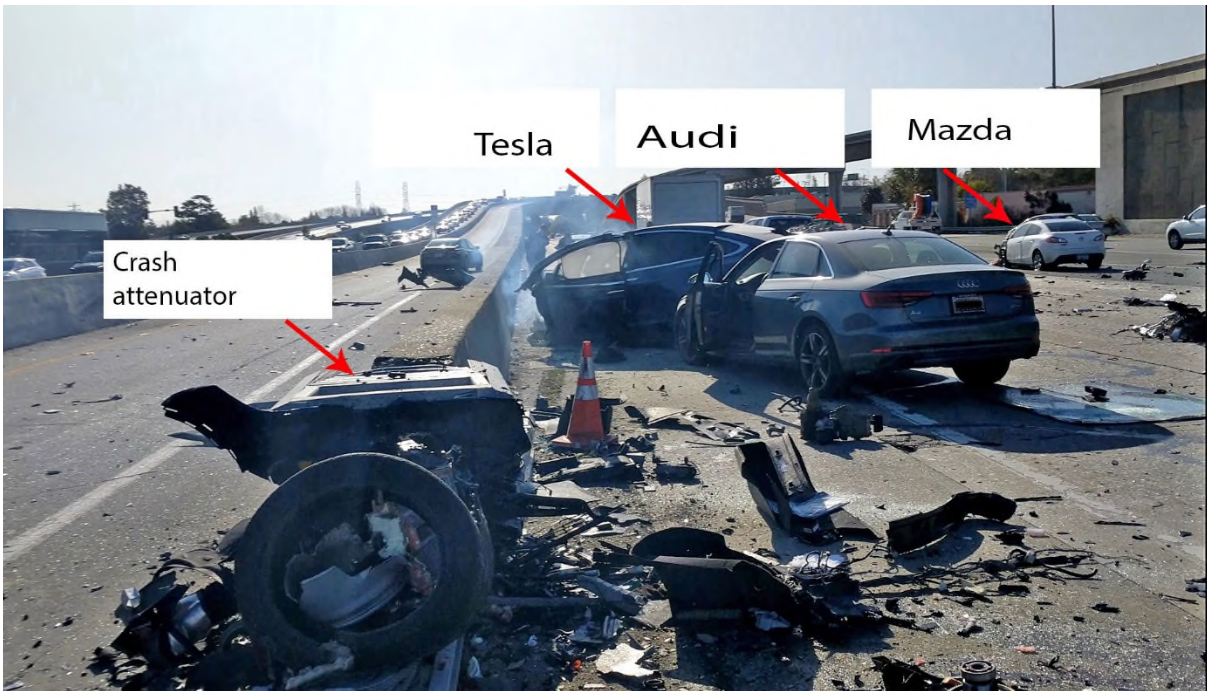
Machine learning moves AI into a more interactive mode, where the program learns from the data, rather than being fed rules. Bayesian Adaptive Probability is a very powerful mathematical technique described by the rev. Thomas Bayes and published posthumously in 1763. Bayesian Adaptive Probability 'learns' from events building up a probability score of something occurring again based on prior observation. It is a very powerful technique that is widely used in computing today. Many programs today monitor your use and preferences and try to anticipate what you want to do next. Statistical Probability does the same, but on a static data set rather than dynamically.

A good example of this is Statistical Machine Translation (SMT), where large corpora are analysed to provide word and phrase based equivalent tables between source and target segments. These alignments are then used to try and 'translate' new source text into the target language on a word or phrase basis. This approach works best with language pairs that have similar morphology and word order such as English and French, or English and Spanish. SMT produces rather poor results where the language pair in question differs significantly in terms of word order and morphology as it is unable to account for grammatical requirements for morphologically rich languages.

Neural Networks are the latest and most inspiring development in AI. Using vast amounts of data, be it chess games, poker tournaments or bilingual corpora, Neural Networks will crunch through looking for and learning new patterns. Unlike the previous techniques discussed, Neural Networks are a closed system and we have no insight on what and how is being learned. The outcomes can be very exciting, like winning a Go tournament with Korean Grand Masters, or poker tournaments, through to Neural Machine Translation (NMT). As exceptional as the results may appear, nevertheless it is important to note the limitations:

1. Neural Networks require vast amounts of very good quality data. Too little data, or data of poor quality will produce bad results.
2. The actual processing power required can be enormous. IBM's Watson AI system requires around 90,000 Watts. An adult male human being requires around 80 Watts.
3. AI systems can only operate on one single topic at a time.

Attempts at autonomous driving cars have shown that real world driving is a whole new ball game. Coping with 80% of normal driving situations and conditions is doable. The remaining 20% poses considerable problems, especially in challenging circumstances such as snow or torrential rainfall, but also concerning something that comes totally naturally to human being but is very challenging to AI: theory of mind. When we drive, we are constantly observing, mostly subconsciously, the actions of other drivers and act accordingly. If someone is driving too aggressively or too timidly we adjust our driving to take these factors into account, or to heighten our awareness to potential problems that may arise, based on prior experience Theory of mind is an inherent human trait that is developed from childbirth and is one of the things that characterizes human interaction. It is not something that you can program for effectively or learn via a neural network. Bridging that remaining 20% gap in autonomous driving may yet prove too challenging to fully achieve.



80% rule: 20% fail results

Being Human

The next stage of our journey into the realm of AI takes into how we, as human beings, relate to our environment and what is language. This is not as simple as it first appears. Evolution has, at it always does, made us more or less fit for purpose. That purpose, in biological terms is to pass on our DNA, or rather a 50% random portion of it to our offspring. This does not mean that we are total masters of our environment, but rather we are made to think we are.

All our senses and perception are honed from birth to allow us to make, more or less, sense of our surroundings and interaction with the real world. Extensive scientific investigation has proved that in reality our contact with reality is somewhat tenuous. We learn how to filter only those aspects of the world around us that are pertinent to our wellbeing and existence.

Take sight: we actually see everything upside down – the eye is a lens that inverts the view of the real world. The brain merely adjusts this automatically, learned from infancy. The human eye has a lot of flaws that would not be acceptable in a camera. We can only see in high resolution in a very small area in the centre of our vision, called the fovea. We have a blind spot where your optic nerve meets up with your retina. You have to move your eyes around a scene not only to take in more information but to correct for these imperfections in your visual system. The eye does not take in a scene in one movie camera type swoop: every 4 seconds it executes seemingly random darting like movements, where it tries to fill in more detail. This is completely subconscious: we are not even aware of it. Our sight is in many ways an illusion designed to allow us to cope with the reality in front of us: much of the detail is made up by the brain, mainly from memory. Magicians make very good use of this fact. The same limitations are apparent with most of our senses. Compared to dogs our sense of smell is incredibly poor, but as with everything else it is sufficient for our existence.

The period of brain development from infancy through to puberty are key in our ability to make sense of our environment, for all of our senses. Someone blind from birth would not be able to make sense of the world of sight if, through surgery, they were given the ability to see later in life:

there is just too much chaos and overload to be able to cope with the rush of images and light. In fact, there is well documented evidence for this. The same goes for sound and even more for speech. Infancy is the period when we learn to adapt and distinguish the key aspects of our surroundings. This period is also a key feedback loop for brain development: it is stimulated and adapts as it learns. At puberty brain development ceases and learning takes on different characteristics. Infancy is a vital period where we come to terms with what we perceive to be the real world and how we can cope with it.

Over the past 50 years we have learned a great deal about the human brain, as well as the brain of other creatures. Arguably the human brain is one of the most marvellous things in the universe. Its complexity and functioning are now closer to being understood. The mathematics underpinning brain activity are truly awe inspiring and very efficient. We even have a better grasp of difficult topics such as consciousness. We have also worked out key aspects of the mathematical basis of key aspects of the brain such as how we can instantly recognize given objects, merely from an outline.

Our interaction with the real world, from waking up to making a meal to driving to work, all entail a detailed internal model of the world and how it functions, within the limitation of our senses discussed previously. These have been learned from infancy and honed as we grow up. Learning, in a human context, entails laying down the neural pathways that are reinforced over time that allow for this to happen. Our brains interpret the world through our senses and build an effective model of reality as we perceive it. We exist within the context of our experience of the world. We learn by recognizing patterns and remember them for future use.

We also fool ourselves into thinking that we are rational and make decisions based on facts. Nothing could be further from the truth. Our social interactions are accompanied by a flood of chemicals that stimulate, please or annoy us. We are a walking chemical factory of enzymes, hormones and various bioactive compounds. Being human is a very complex process from the chemical point of view.

Language

Human beings are social primates: we share 95% of our DNA with our closest ape cousins. One of the main differences, that account for some of the remaining 5% is a gene called FOXP2. This is the language gene. Apes groom, we talk. Solitary confinement is one of the worst psychological forms of torture for humans. As social animals, we need each other. Language can only be understood from a human perspective: it is not some kind of scientific or logical notation. Language is typically full of anomalies, exceptions and much repetition. Language is human, and along with most human characteristics it is never completely rational.

The best way to describe Language is to compare it to clay in terms of pottery. Clay represents the potential. The final fired ‘pot’ represents Language and individual vessels can have different shapes and sizes. Language can take various forms, from sign language to the whole variety of human languages spoken today. For spoken language, the ability to force air through our voice box and form different sounds using our tongue and lips allows us to talk. Interestingly it takes many years from infancy to around 7 years old to master the correct tones for a given language. Languages vary quite considerably in terms of pitch and sound, which is why learning to speak in a different language leaves most of us with an accent. After puberty, it is much more difficult to articulate fluently in another language.

The key elements of Language are things (nouns) and actions (verbs), followed by the plethora of additional parts of speech required to provide more detailed meaning. All languages have these. The actual grammar of a given languages can vary considerably, but they must be able to convey the same meaning in human forms. For example, you can have languages with no verb tenses: you just have to provide sufficient context to establish who did what to who when. You can also have languages with no plural form; again, you just have to provide sufficient context.

And then came Computers

The basic architecture of all mainstream computers, be they smartphones to supercomputers, was laid down more than 70 years ago by, among others, John von Neumann and Alan Turing. The basics remain unchanged: processors have grown immensely more powerful, doubling in performance every 24 months on average. Memory and storage has also grown at an incredible rate, but the basics remain the same. You can compare the basic inherent intelligence of our processors to that of a tape worm. A tape worm inhabits a warm dark space and has absolutely no knowledge of the working of the outside world. The same goes for our computers.

The only 'intelligence' that we can gain from the current generation of processors is via computer programs. What computers can do, as opposed to human beings, is to sift through unimaginable scales of data and look for patterns. This data can be sequences of moves in a game, or vast amounts of text, looking for patterns. As computers have no knowledge or experience of the outside world their view is strictly limited to the task at hand. There is a great deal of confusion regarding AI. Alan Turing defined the 'Turing Test' in 1950 as "a computer's ability to exhibit intelligent behaviour equivalent to, or indistinguishable from, that of a human". This famous test does not necessarily provide a basis for intelligence as Alan Searle pointed out in his "Chinese room' thought mind experiment in 1980 – although the interaction might appear to have a surface form of intelligence, computers are inherently incapable of 'understanding' which is a core issue in general intelligence and in navigating the real world.

The current generation of computers differs substantially from the functioning of the human brain or the brains of mammals and other creatures. The basic computer chip has no inherent intelligence. It is merely a tool for processing data. It has no knowledge of the physical world and only executes programs that are fed to it. In many respects, it is like a tape worm: both live in warm dark places totally isolated from the outside world. What computers can do though is process vast amounts of data – something which is outside the practical scope of individual human beings.

The von Neumann architecture has served us extremely well, but it has its limitations: it is excellent for databases and spreadsheets and we can do some amazing things by crunching very large amounts of data for an extended period to learn new patterns. There are many built in legitimations though that limit the potential scope for the current design of computers.

The leap to a general purpose intelligent and autonomous computing device will require a substantial rethink. Much very good work is already in progress based on a more detailed understanding of the mathematics and detailed working of the human and generally mammal brains. Neocortical computing is an exciting new field that is still in its infancy but promises a totally new approach to computing that may one day lead to truly autonomous intelligent devices. It attempts to mimic closely the function of the neocortex in human brains and the neurological pathways that link everything together to produce a truly autonomous intelligent device with scalability and very low power requirements.

Expert Systems

Expert systems apply the same rules to a problem that an expert would employ. There are two ways of doing this. The first 'Knowledge based' method encodes the 'expert system' as a series of program functions: for example working out the optimal layout for a table, taking into account the number of rows and columns and the proportion of header text to the column text. The net effect is what an experience typographer would have done: we have encapsulated the table layout skills of a typographer in order to automate the table layout process.

The second approach regarding expert systems is to encapsulate the whole domain information for a given topic into a knowledge base as a set of rules: if 'X' then do 'Y'. The knowledge base is then used to drive the decision-making process rather than specially coded algorithms as in the knowledge-based approach. An example of this is a general-purpose diagnostic system, where information is obtained about a symptom that will lead to further questions that narrow down the possible problem until a definitive cause is established beyond all probability.

Both types of expert systems are based on existing domain knowledge, and both qualify as AI in meeting the core three 'intelligence' criteria detailed above: pattern recognition, storing of pattern information and using the stored patterns to predict a course of action.

Machine Learning

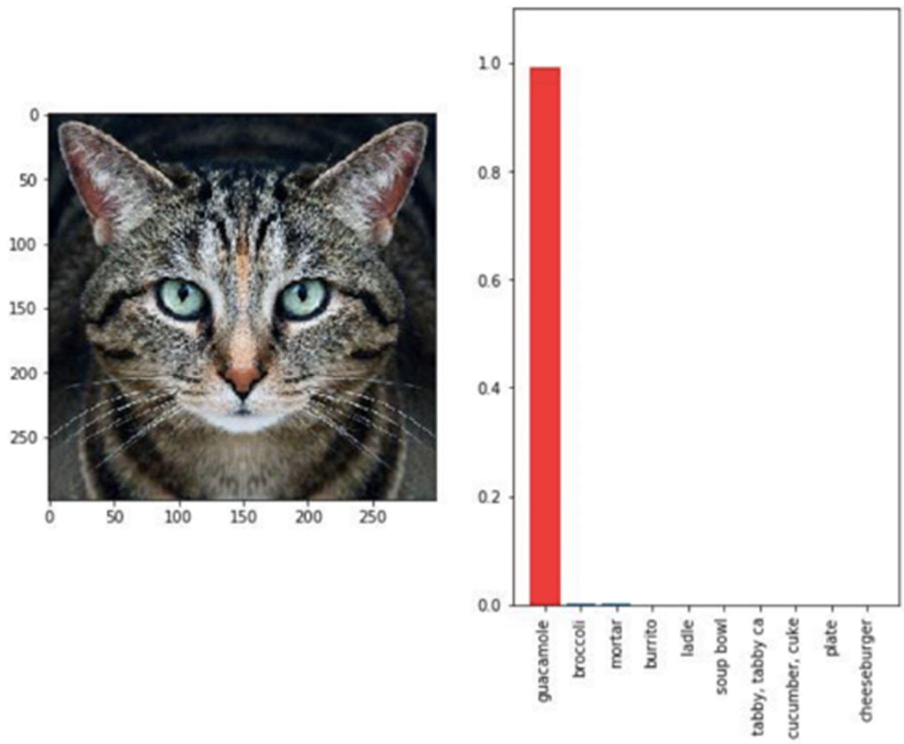
Whereas in expert systems we capture patterns a priori, with machine learning we let the system work out what the patterns are and to then store and use those patterns to solve problems.

Bayesian based ML works out the probability of an occurrence by observing a series of events and storing that information for future use. Bayesian probability is also known as adaptive probability, where the probability of an event increases with the frequency of its occurrence. The more frequently something happens, then the probability of it repeating increases. Bayesian ML provides a detailed audit trail that can be observed and verified and is very fast in operation. Bayesian systems can also be linked in a network known as a Bayesian Belief Network. Bayesian based ML relies on a simple, but very powerful mathematical probability function described by Rev. T. Bayes (1701 - 1761).

Neural Network based ML uses a very different approach. Neural Networks rely on an extremely large amount of data to produce an output engine that can then be used to classify, for instance images, or for machine translation. The data is fed into the system which analyses the data according to pre-set criteria. There is no control over the decision process: the system is left to its own devices and does not provide any information as to why it has made specific decisions: it is effectively a 'black box' that cannot be tuned in any way.

The quality and makeup of the input data is critical. Neural Networks will reflect any bias inherent in the input data and any errors will adversely affect the output. The other distinguishing feature of the neural network approach is the very long run times and computational requirements compared with other systems mentioned. In many cases neural networks will provide similar decisions to expert systems, but have the benefit of being more precise.

Neural Networks can be 'brittle' depending on the amount and quality of the training data. Small variations can produce unexpected results as evidenced in the following image misclassification from one of the main image identification engines:



The other problem with Neural Networks concerns updating the engine with new data. As training times can be measured in weeks, engines cannot be easily updated with new data.

AI in Localization

It comes as a surprise that most Computer Assisted Translation (CAT) tools have been using AI for many years in one form or another. Translation Memory, Terminology Extraction, Machine Translation are all aspects of pattern matching and qualify as AI based on the standard definition of intelligence. In addition, Statistical Machine Translation and Neural Machine Translation.

There are many aspects of Localization that can and do benefit from AI. Many of these aspects work in conjunction with significant lexical resources such as bilingual lexicons as well as Part Of Speech (POS) analysers:

1. Intelligent corpus alignment that 'learns' as it aligns based on lexical and POS resources
2. Terminology extraction using POS analysis
3. Bilingual terminology extraction using multi-lingual POS analysis
4. Automated workflows that route data to reduce workload, such as automatically defining which project and language pairs require or do not require proofreading or additional checks
5. Automatic transfer of inline elements from source to target segments
6. Fuzzy match completion using bilingual lexicons and POS analysis
7. Dynamically 'learning' translation from the translator as he/she translates, building a Machine Translation (MT) engine on the fly.

Neural Networks have shown extremely good results for Machine Translation (MT). The output for Neural MT (NMT) is much more fluent than that produced by Statistical MT (SMT):

German Source: Probleme kann man niemals mit derselben Denkweise lösen, durch die sie entstanden sind.

English Translation:

SMT: *No problem can be solved from the same consciousness that they have arisen.*

NMT: *Problems can never be solved with the same way of thinking that caused them.*

The fluency of the output of NMT far exceeds that of SMT. In addition SMT is only really effective for languages that have a relatively primitive morphology and similar word order grammar. The downside of NMT is that an order of magnitude greater amount of training data is required and it cannot cope at all well with new words in the source language that were not present in the training data: it either repeats the preceding word or inserts a 'random' word to complete the segment:

French Source: Les adolescents japonais aiment les jeux vidéos.

SMT: *The adolescents japanese love electronic vidéos.*

NMT: *Japanese teenagers are interested in fashion.*

The fluency of NMT output poses another problem for post-editors: attention can wane when the output looks so good and it is much more difficult to spot errors when they do occur.

References

Jeff Hawkins and Sandra Blakeslee, On Intelligence, 2005, ISBN-10: 0805078533

Lesa Moné,, Artificial Intelligence - Expert Systems, 2011, https://blog.leanix.net/en/artificial-intelligence-expert-systems

M.K. Bergman, Knowledge-based Artificial Intelligence, , http://www.mkbergman.com/1816/knowledge-based-artificial-intelligence/

Crina Grosan, Ajith Abraham, Rule-Based Expert Systems, pp 149-185, Springer-Verlag 2011. ISBN 978-3-642-21003-7

John McCarthy, Artificial Intelligence (AI) vs. Machine Learning vs. Deep Learning, 2019, https://skymind.ai/wiki/ai-vs-machine-learning-vs-deep-learning

M. Kukacka, Bayesian Methods in Artificial Intelligence, WDS'10 Proceedings of Contributed Papers, Part I, 25–30, 2010. ISBN 978-80-7378-139-2 © MATFYZPRESS

Toma´s Mikolov et ales, Recurrent neural network based language model, Interspeech 2010, https://www.isca-speech.org/archive/archive_papers/interspeech_2010/i10_1045.pdf