Neocortical Computing: Next Generation Machine Translation

Andrzej Zydroń MBCS
CTO XTM International Ltd.
azydron@xtm-intl.com

Abstract

We have reached the theoretical limits of what can be achieved through the application of Statistical, Rule based and Transfer based machine translation technology. The limits are those imposed by the Turing architecture which is what we are currently restricted to. The start of the 21st century has seen significant theoretical advances in the domain of human intelligence and its mechanisms and underpinnings.

1 Introduction

To date we have relied on the basic computing architecture as laid out by Alan Turing during the late 1940s. Little in essence has changed concerning the basic framework, encompassing a CPU, processing instructions and volatile and non-volatile memory stores. It has served us quite well and we can all see the benefits around us in our daily lives from automatic ticket machines to tablets. Nevertheless this approach has many practical limitations when it comes to trying to address the complex world of intelligence and that uniquely human and idiosyncratic method of verbal communication that we call language. Alan Turing postulated the 'Turing test': a test of a computing device's ability to exhibit intelligent behaviour, equivalent to or indistinguishable from that of a human being. We have recently seen examples of systems that purport to have passed this test (IBM's Deep Blue in terms of chess and Jeopardy).

2 Why can't Computers do That?

The eminent philosopher, John Searle in his famous 'Chinese Room' thought experiment, showed the limitations of the Turing test. There are very simple everyday things that we take for granted that pose almost insurmountable problems for the current generation of computer:

- Recognize from just a few lines the outline or a dog or a cat.
- Recognize that a cartoon cat is a cat.
- Understand free flowing conversations.
- Learn how to walk.
- Walk freely across a rubble.
- Find a style on a footpath and climb over it.
- Catch a ball that is thrown up in the air freely.
- Learn from experience.

A two year old dog can run, jump and catch a ball or stick thrown up in the air. Yes, you can achieve some of these things with an enormous amount of brute force and many man years of programming, but these will normally be limited to a single detailed application and

Proceedings of the 37th Conference Translating and the Computer, pages 102–105,
London, UK, November 26-27, 2015. ©2015 AsLing

nothing else. Anyone who has watched the latest DARPA robot trials will appreciate how difficult it is to program a robot to do the simplest of special tasks without very comical results, and this from the very best engineering and university teams in the world.

3 The AI Brick Wall

When it comes to trying to delve deeper into the realms of artificial intelligence that things start to unravel. The Turing machine has all of the hallmarks of the 'if your only tool is a hammer, then all problems looks like nails' syndrome. Various attempts have been made to create 'intelligent' programs, but in reality all we end up is with so called 'expert systems' that encapsulate where possible the standard rules that are applied to a given problem by an experienced practitioner. The great promise of AI in the 1980s soon evaporated. The problem still remains: maybe the hammer is not the right tool.

Subsequently we have tried to solve the problem by brute force, as with IBM's Deep Blue and Jeopardy engines or by clever mathematics, but in the end these systems lack the basic ingredient: they do not understand. You can extrapolate great insights from big data and we can have quite a degree of success with treating language translation as a piece of cryptography, but there are real finite limits to what can be achieved and the best systems still require a great deal of manual 'tuning'.

4 The Limits of Machine Translation

Machine translation made a great leap forward at the start of this century thanks to the seminal work by IBM researchers working on the benefits that exploitation of Big Data in the form of massive scale aligned corpora could have on treating translation as a cryptographic problem. Further advances were made thanks to funding from the European Union in the form of the moses project. Philippe Koehn, Franz Josef Och and Daniel Marcu were the main researchers that worked on the first production SMT (Statistical Machine Translation) systems. There are currently dozens of SMT systems available online from Google Translation, Microsoft Translator, Asia Online and many more commercial or academic offerings. A well-trained engine can improve translator productivity by around 10 – 25 percent.

Assuming that the main goal is to improve translator productivity, rather than providing a 'gist' translation there are some significant limitations to SMT:

- Larger amounts of data do not result in improvements to the performance past an optimal amount. This is caused by the problem of 'noise' arising around polysemy, which is ever present in human language.
- The demands of morphology can distort the resultant decoded text to a very large degree.
- Out of vocabulary words, which the decoder has not previously encountered.
- Differences in context between the training material and the text being processed can have a great bearing on the accuracy of the output
- Most engines require manual 'tuning' to produce the best results
- Engines cannot be tuned effectively on the fly.

In the final analysis the quality of the output is equal to the amount of effort that has gone into training and tuning the particular engine. It is as if the first law of thermodynamics holds sway: in order to provide an improvement in translator productivity, you need to expend appropriate resources in training and tuning the SMT engine.

5 Neocortical Approach

What was required was a completely different approach: how do you define intelligence, how can you quantify it, and how do you build systems that can be both truly intelligent and learn by themselves. These questions occupied Jeff Hawkins when he studied at both Cornell and Berkeley University. A gifted computer engineer, Jeff Hawkins was also the technical architect behind the Palm Pilot and Treo devices. The problem that Hawkins discovered was that there were no good or bad theories about intelligence and how the neocortex, which is common to all mammals, actually functions to produce coherent actions, and most importantly how it actually works.

Jeff Hawkins published his seminal work 'On Intelligence' in 2005. 'On intelligence lays out the fundamental theoretical mechanism behind the way in which all mammalian brains function. The neocortex is fundamentally different in all respects from the Turing architecture.

6 Pattern matching Machines

The neocortex in human beings is roughly the size of a napkin and is made up of six layers, each the thickness of a standard business card. It is folded up into the characteristic form than we see on the outside of the brain in order to fit into the cranial cavity.

The human brain is very slow compared to that of the modern CPU. At best it can manage around 100 discreet operations per second (on a good day when you are in your early twenties – it is downhill all the way after then). We actually use two approaches: a low cost 'slow' brain which we use in normal everyday instances like walking, making teas etc. and a high maintenance brain which we use when concentrating on a particular task such as counting, or working out a detailed problem. The two do not mix well: try walking backwards and counting down from 100 to zero. In comparison the current tablet or mobile phone processor can manage 3 billion operations per second. What the mammalian brain does have though is trillions of connections.

The essence of learning and understanding lie in the way information is stored and retrieved in the neocortex. All animals are in essence pattern matching machines. We exploit patterns in nature, the seasons, day and night to exist and multiply.

How can we, in the blink of an eye, recognize someone from a distance, just by their demeanour or gait. How can we tell, without thinking out it, when we see any form of dog, from the vast variety or actual breeds, through to a cartoon dog, that it is a dog.

7 Invariant form and Hierarchical Structure

At the core of the way that the neocortex works is the concept known as 'invariant form'. The mammalian brain's main mechanism to pattern matching is to categorize. Categorization depends on associating what is known as an invariant form with an item that it is observing. An example of an invariant form is 'horse'. Under this concept are grouped all instance of 'horse'. This is how the brain copes with the rich and varied reality that surrounds us. Categorization is assisted by the six layers of the neocortex. The structure of the layers allow for the almost instantaneous recognition of an object as such. From the computer science point of view the six layers present a bitmap gate and simple and very effective 'and' and 'or' operations allow for the recognition process.

8 Synapse Connections

The various aspects of the neocortex are all interconnected via synapses which bind the main parts together. What the brain lacks in speed it makes up for with over one trillion connections. These connections are key to how the brain 'learns' to cope with the external world. It is very effective, and the result of billions of years of adaptation and .

9 Language

The nature of language is a typical adaptation of the brain to the problem of verbal communication for homo sapiens. Idiosyncratic, full of inconsistency and illogical contractions, incredibly varied and messy, language has long been a 'bad fit' for current state of computational methods and for current Turing based computer architectures. The human, messy nature of language defies the simple algorithmic approach of computer science, which is more at attuned to databases and data analytics than to the more complex issues that deal with everyday reality.

10 Conclusion

The work of Jeff Hawkins has provided the basis for a new approach for the next generation of computing devices. In order to build truly 'intelligent' machines we need a completely different approach. Neural networks and Bayesian Belief Networks, which have provided some solace in mapping the messy entropic nature of reality onto our current Turing approach to computing, nevertheless have serious practical limitations and in reality constitute a dead end. In essence we are currently armed with a hammer while trying to solve a problem that requires highly complex and adaptable machine tools.

The neocortical approach has generated a lot of interest, both from the hardware and software points of view. Both Qualcomm and IBM have started to lay down neocortex based silicon. On the software side there have also been some very interesting developments. Vienna based cortical.io have built a natural language processing engine based on neocortical concepts and Jeff Hawkins has set up Numenta, a software company to build self learning programs using his latest theories.

The coming decades will see some very interesting advances in terms of language processing and translation based on the neocortical approach. Numenta and cortical.io are showing the way.

References

Jeff Hawkins, Sandra Blakeslee. 2005. *On Intelligence*, Times Books, Henry Hold and Company. ISBN 0-8050-7456-2

Jeff Hawkins: *How brain science will change computing (Speech).* TED 2003. Retrieved 2014-05-09. http://www.ted.com/talks/jeff_hawkins_on_how_brain_science_will_change_computing.

Jeff Hawkins, 2012-10-19, *Computing Like the Brain*, InfoQ http://www.infoq.com/presentations/Brain-Computing/.