Quantum Cognition Machine Learning
AI Needs Quantum

Kharen Musaelian*, President and Chief Science Officer
Alexander Abanov, Jeffrey Berger, Luca Candelori
Vahagn Kirakosyan, Ryan Samson, James Smith, Dario Villani

Qognitive, Inc.

March 1, 2024

1 Introduction

The last decade has seen massive improvements in machine learning techniques that hold promise for breakthrough advances in productivity and technology. Parallel to this progress, our way of discussing new techniques and their shortcomings has evolved to anthropomorphic tones with proliferation of terms like attention, intelligence, hallucination, bias, and insight. At the same time, there is a growing understanding that something fundamental is missing in the current framework of generative AI and statistical machine learning in general [1, 2]. In this paper, we outline why statistical machine learning based on classical probability theory is fundamentally inadequate for the tasks of attention, relevance realization, abstraction and concept generation, and situational awareness – tasks that humans and animals excel at. We argue that a machine learning approach based on the ideas of quantum cognition ([3] and references therein) is needed to bridge the gap between where we are and where we wish to be. We also describe our implementation of the quantum cognition machine learning paradigm at Qognitive.

2 Shortcomings of Classical ML

Whether implicitly or explicitly, machine learning has always been about learning a joint probability distribution of preprocessed features, whose relevance and significance is ascertained by a human understanding of the domain. Here, the joint probability distribution is understood in the classical Kolmogorov sense, based on classical sets, whereby inference is achieved through application of Bayes' rule. The fundamental problem with this classical probabilistic description is that its complexity grows exponentially with the number of features.

*Primary Investigator (kharen@qognitive.io), remaining authors presented in alphabetical order

For example, a probability distribution over N binary features corresponds to a vector of size $2^N - 1$[1]. This also leads to an exponential requirement for the amount of data needed to learn this distribution statistically. In practice, statistical analysis is only made possible by a careful choice of relevant variables and overall dimensionality reduction based on human insight and understanding. The basis of this human insight and understanding is arguably not statistical [4, 5], otherwise one falls into a circular explanation [6]. This elicits the fact that even the most advanced AI to date has shown remarkable lack of common sense and understanding [1, 2, 4, 5, 7].

Intimately related to this curse of dimensionality[2] is the problem of representing concepts through classical sets. A good illustration of this problem is that the number of non-empty subsets of a set with N elements is also $2^N - 1$. For the case of infinite countable sets, such as the set of Turing machines or statements of a formal number theory, the set of subsets represents a non-countable infinity, which is at the core of the proof of Church-Turing[8, 9, 10] and Gödel's theorems[11][3] [4, 12]. As an aside, this very realization led Alan Turing to conclude that "if a machine is expected to be infallible, it cannot also be intelligent [12]." Consequently, generating new concepts as subsets of existing ones leads to the same curse of dimensionality.

For example, consider GICS (General Industry Classification Standard) as it has been applied to Iron Mountain (IRM). Iron Mountain is a company specializing in data storage and security. Before 2020, IRM was classified as a Commercial and Professional Services company, but in 2020, it was reclassified as an Equity REIT (Real Estate Investment Trust) purely for tax purposes. IRM is a REIT and a commercial services company at the same time. One could fault the GICS system for not having sufficiently well separable categories, but the fact is that any classification system encounters the same problem – classical sets allow their elements to be only 100% in or 100% out. If we try to solve this problem by subdividing every commercial services company into REITs and non-REITs, we end up with too many buckets, making generalization impossible. We could also consider GICS assignment as a "noisy" observation and apply Bayes' rule to create a "probabilistic mix" of Commercial and Professional Services and REIT, but the problem is that depending on the context, IRM is 100% REIT or 100% Commercial and Professional Services. Bayesian probabilistic mix of the two does not make sense, and the observation is not at all "noisy".

Another difficulty of classification within the classical statistical and set the-

[1] For a sense of scale, 2^{300} is more than the number of atoms in the universe.

[2] The curse of dimensionality in statistics refers to exponential growth in the amount of data needed to achieve the same accuracy of parameter estimation as a function of the number of parameters.

[3] Godel's theorem is a result in mathematical logic that states that any self-consistent formal number theory contains true statements that are unprovable within the formalism of the number theory. Closely related to Godel's theorem is the Church-Turing theorem that shows that there is no Turing machine (classical computer program) that can determine if any given Turing machine given a particular input parameter will ever halt. The proofs of both theorems rely on the fact that both the set of all statements of a formal number theory and the set of all Turing machines are countable infinite sets, i.e. sets that can be enumerated by whole numbers. In contrast, the set of all subsets of a countable infinite set is not countable, i.e. is a more powerful set. The fact that human mathematicians can see the truth value of the Godel statements has led Lucas and Penrose to argue that human cognition cannot be simulated by a Turing machine.

oretical framework is related to clustering based on a form of distance. Firstly, clustering algorithms notoriously suffer from the curse of dimensionality [13]. More subtly, one notes that human language is profoundly metaphorical, with analogies entirely context dependent. Consider the sentence "Snow blankets the ground" (due to Julian Jaynes [14]). Here, "snow" is compared to a "blanket". Any sensible clustering algorithm would consider "snow" and "blanket" as very distant objects, and in fact, any two objects are similar in infinite number of ways, most of them strikingly irrelevant ("humans and chairs are both smaller than the sun"), and different in infinite number of ways [5]. What makes a metaphor work is our ability to abstract away all the infinite number of ways that the two elements of a metaphor are different and bring into crisp focus the way they are similar, something that is mathematically impossible within the confines of classical sets.

This difficulty of creating abstraction within a classical Bayesian framework is also illustrated by the fact that in classical probability any uncertainty is associated with entropy. Thus, in the case of IRM, a Bayesian approach would have to assume that the observation of IRM as a REIT is somehow "noisy" and is associated with a probability of being "wrong". A surprising answer to these problems lies in using quantum, rather than classical probabilities.

3 Quantum cognition

The first idea of quantum cognition emerged from the works of Aerts et al [15], Khrennikov [16], and Busemayer et al [3] (see [17] for a recent survey). In these works, it is posited that the state of mind is formally given by a quantum state, i.e., a vector in a Hilbert space, and all questions that can be answered within that state of mind are represented as operators in that Hilbert space. This way of representing the state of mind, called Hilbert Space Models, turned out to be remarkably successful at explaining our cognitive fallacies. Within cognitive science, this discovery was treated as a mere curiosity. However, at Qognitive, we have discovered that representing data as a vector in a Hilbert space with observables represented by operators (matrices), leads to a logarithmic reduction in the complexity of representation. This dramatic economy of representation explains why evolution would select quantum cognition over classical statistical learning. As living creatures, we do not encounter the world as well-structured data relevant to a task at hand. Instead, we are confronted by a barrage of unstructured inputs that need to be made sense of, while focusing on what is important and drawing conclusions by abstracting away what is irrelevant. Note that neither the pioneers of Quantum Cognition, nor we, take any position on whether our brains achieve this form of representation by being physical quantum devices [3, 18]. What we claim is that this formalism of quantum probabilities describes our state of mind and our representation of data in a way that achieves logarithmic reduction in complexity and makes non-statistical data inference possible.

One dramatic consequence of this form of data representation and inference is that various observables are represented by non-commuting operators. Therefore, like position and momentum of an electron in quantum mechanics, observables are subject to the uncertainty principle – in general, there is no quantum state that corresponds to an exact observation of all the data. The cost of the logarithmic economy of representation is that the encoding of a large number of features always involves a loss of precision. Another cost is what led to the idea of Quantum Cognition in the first place – the cognitive fallacies that represent violations of Kolmogorov set rules, such as conjunction and disjunction fallacies [3].

Representing data as a quantum state allows to see IRM (our example company from sec:2) as a Schrödinger's Cat – a superposition of states corresponding to various business models. More broadly, we argue that the uncertainty principle inherent in quantum formalism is not a bug, but a feature. Recall that in classical statistics any uncertainty is associated with non-zero entropy. Now, consider the Bell state of two spins given by a superposition of spin-up spin-down and spin-down spin-up $\frac{1}{\sqrt{2}}(|\uparrow\downarrow\rangle - |\downarrow\uparrow\rangle)$. This is a pure quantum state, i.e., it is associated with zero entropy. However, considered separately, both spins are in a state of maximum uncertainty – in quantum theory, one can and necessarily has uncertainty in the absence of any entropy. This is what abstraction is! For example, consider the statement "1+1=2". Numbers are meant to refer to objects whose quantity they represent. However, the meaning of the statement "1+1=2" is fundamentally different, and much crisper, than the statement "one apple plus one apple equals two apples". The latter statement invites all sorts of questions of whether the two apples are in fact equivalent, whether an apple could be added to an orange, and so on. Abstraction always requires a degree of uncertainty, which is the flip side of generalization.

Importantly, just as quantum mechanics becomes classical in the limit where Planck's constant[4] becomes 0 (when commutators also tend to zero), quantum representation of data becomes classical in the limit of high dimensional Hilbert space, where all operators can be rendered commutative. On the other hand, no amount of complexity can turn classical formalism into quantum.

4 Quantum Cognition Machine Learning

Using the insight of the pioneers of Quantum Cognition [3], we found a practical way to implement a new paradigm of machine learning, which we call Quantum Cognition Machine Learning (QCML). While substantially deviating from the original models, it broadly follows the principles of quantum cognition outlined above. In the context of QCML, training involves learning the operators and encoding data in quantum states. All the correlations and higher order relationships within data are thus captured within the form of the operators and their rules of commutation. Crucially, context dependent relevance of features, concept generation, and abstraction arise naturally in this framework.

Additionally, QCML's economy of representation can lead to drastic reduction of computation costs incurred by brute force solutions to AI problems [2].

Our QCML framework has been implemented on classical computers and benchmarked against state-of-the-art machine learning methods in a variety of settings as a proof of concept. Examples include MNIST [19], COIL20 [20], PenDigits [21], and a variety of time series benchmarks [22].

[4] Planck's constant, often written as h or $\hbar$, is one of the fundamental constants of quantum mechanics. If this constant is set to 0 then quantum behavior disappears and classical behavior is recovered.

Yet, our greatest advantage may lie in the settings that are not even covered by current machine learning methods, in particular, situations where the number of features far outstrips the amount of data you would need to learn a statistical model. A good example is epigenetics – the study of gene expression. Epigenetics holds tremendous promise for cancer diagnostics, care, and prevention. Yet, the current data inference methods only allow us to scratch the surface [23] – a typical study focuses on a few genes at a time, leaving the vast majority of data untapped. Consider that the human genome has about 28 million methylation sites, and this data is available to be linked with variables, such as diagnoses, response to various therapies, etc.

Other examples where QCML can be applied are finance, advertisement technology, large language models, robotics, defense, materials, etc.

5 QCML and quantum computing

Mathematically, QCML is a simulation of a quantum system, and as such it has a natural extension to quantum computing. Qognitive is building partnerships to explore a variety of physical quantum implementations of QCML on universal quantum computers and specialized quantum devices. Quantum computing extensions of QCML hold the potential to achieve true quantum advantage in machine learning by exponentially enlarging the Hilbert space. Here, the process of training corresponds to learning how to configure a physical process for a given data point. We encode the data of the problem into a Hamiltonian and through finding the ground state of this Hamiltonian we encode the data into the quantum state, a method which fits naturally into quantum systems. Our encoding method is distinct from existing quantum methods such as amplitude or angular encoding, which seek to directly encode a bytestring into a quantum state[24]. Intriguingly, executing QCML on native quantum hardware would create the first example of a non-Turing machine [25] that could overcome the well-known limitations of Turing machines [4, 12].

6 Conclusion

Qognitive discovered and implemented a new machine learning paradigm using principles of quantum cognition. The new paradigm achieves logarithmic economy of data representation and is capable of non-statistical learning, including relevance realization, attention, abstraction and generalization, and concept generation. We continue finding new ways to improve QCML, combine it with existing machine learning methods, and developing methods that are only applicable within the quantum paradigm.

As we improve our methods and understanding and optimize our digital computer implementation, we are developing concrete business applications. In particular, we are already generating financial forecasts that are currently used in asset management.

We are also working on building a bridge to quantum computing, to which QCML has a natural affinity.

7 Acknowledgements

The author would like to ackwnowledge the support and contributions of the following members of the Qognitive advisory board: Mario Bondioli, Tomaso Poggio, and Nicola Marzari.

References

[1] A. Dawid, Y. LeCun. Introduction to latent variable energy-based models: A path towards autonomous machine intelligence. *Les Houches Summer School Lecture Notes*, 2022.

[2] P. Burchard. Linkedin posts, 2023-2024. https://www.linkedin.com/in/paulburchard.

[3] J. Busemeyer, P. Bruza. *Quantum Models of Cognition and Decision*. Cambridge University Press, 2012.

[4] R. Penrose. *Shadows of the Mind: A Search for the Missing Science of Consciousness*. Oxford University Press, Inc., USA, 1st edition, 1994.

[5] J. Vervaeke. Awakening from the meaning crisis, 2019. https://www.youtube.com/playlist?list=PLND1JCRq8Vuh3f0P5qjrSdb5eC1ZfZwWJ.

[6] J. Vervaeke, Y. Lillicrap, B. Richards. Relevance realization and the emerging framework in cognitive science. *J. Log. Comput.*, 22:79–99, 02 2012.

[7] Y. Choi. The curious case of commonsense intelligence. *Daedalus*, 151:139–155, 2022.

[8] A. M. Turing. Computability and λ-definability. *Journal of Symbolic Logic*, 2(4):153–163, 1937.

[9] A. M. Turing. On computable numbers, with an application to the entscheidungsproblem. *Proceedings of the London Mathematical Society*, s2-42(1):230–265, 1937.

[10] A. Church. A set of postulates for the foundation of logic. *The Annals of Mathematics*, 33(2):346, 1932.

[11] K. Gödel. Über formal unentscheidbare sätze der principia mathematica und verwandter systeme. *Monatshefte für Mathematik und Physik*, 38(1):173–198, 1931.

[12] A. M. Turing. Lecture to the London Mathematical Society, 1947.

[13] I. Assent. Clustering high dimensional data. *WIREs Data Mining and Knowledge Discovery*, 2(4):340–350, 2012.

[14] J. Jaynes. *The origin of consciousness in the breakdown of the bicameral-mind*. Houghton Mifflin, Boston, MA, November 2000.

[15] D. Aerts, S. Aerts. Applications of quantum statistics in psychological studies of decision processes. *Foundations of Science*, 1(1):85–97, 1995.

[16] A. Khrennikov. Quantum-like brain: "interference of minds". *Biosystems*, 84(3):225–241, 2006.

[17] E. M. Pothos, J.R. Busemayer. Quantum cognition. *Annual Review of Psychology*, 73(1):749–778, 2022.

[18] M.P.A. Fisher. Are we quantum computers, or merely clever robots? *International Journal of Modern Physics B*, 31(07):1743001, 2017.

[19] Y. LeCun, C. Cortes. The MNIST database of handwritten digits, 2005.

[20] S. A. Nene, S. K. Nayar, H. Murase. Columbia university image library (COIL-20), 1996.

[21] E. Alpaydin, F. Alimoglu. Pen-based recognition of handwritten digits. UCI Machine Learning Repository, 1998.

[22] A. Das, W. Kong, A. Leach, S. Mathur, R. Sen, R. Yu. Long-term forecasting with TiDE: Time-series dense encoder, 2023.

[23] Q. Zhou, G. Kang, et al. Epigenetic analysis of cell-free dna by fragmentomic profiling. *Proceedings of the National Academy of Sciences*, 119(44):e2209852119, 2022.

[24] M. Schuld, F. Petruccione. *Supervised Learning with Quantum Computers*. Quantum Science and Technology. Springer, 2018.

[25] B.J. MacLennan. Natural computation and non-turing models of computation. *Theor. Comput. Sci.*, 317(1-3):115 – 145, jun 2004.