COMPLEXITY CONTROLLED NEURAL NETWORKS

HIMANSHU PANT

DEPARTMENT OF ELECTRICAL ENGINEERING

INDIAN INSTITUTE OF TECHNOLOGY DELHI

OCTOBER 2021

COMPLEXITY CONTROLLED NEURAL NETWORKS

by

HIMANSHU PANT

Department of Electrical Engineering

Submitted

in fulfilment of the requirements of the degree of Doctor of Philosophy

to the

INDIAN INSTITUTE OF TECHNOLOGY DELHI

OCTOBER 2021

Certificate

This is to certify that the thesis entitled **"Complexity Controlled Neural Nets"**, being submitted by **Himanshu Pant** for the award of the degree of **Doctor of Philosophy** to the Department of Electrical Engineering, Indian Institute of Technology Delhi, is a record of bonafide work done by him under my supervision and guidance. The matter embodied in this thesis has not been submitted to any other University or Institute for the award of any other degree or diploma.

Dr. Jayadeva

Professor

Department of Electrical Engineering,

Indian Institute of Technology Delhi,

Hauz Khas, New Delhi - 110016,

INDIA.

Acknowledgments

I would like to thank my supervisor Prof. Jayadeva, Research Committee members Professors Santanu Chaudhary, I.N. Kar and Amit Kumar; faculty members with whom I worked and published - Professors Suresh Chandra and Amit Bhaya; friends and colleagues Sumit Soman, Mayank Sharma, Pawas Gupta; department staff members Rakesh, Yatindra, Mukesh and Ritwick; my mother,wife, kids and brother for supporting me through this journey.

(Himanshu Pant)

Abstract

The Vapnik–Chervonenkis (VC) dimension of a learning machine is a key concept in the area of model complexity control. The VC dimension is a measure of the capacity of a learning machine. With rapid increases in the availability of training data, and subsequent advent of monolithic deep learning architectures, capacity control of a learning machine has become even more significant. The number of weights in many deep architectures significantly outnumbers the dataset size. The capacity of a machine directly affects its generalizability, due to which architectures with excess capacity tend to over-fit, resulting in poor test performance.

The relationship between the capacity of a learning machine and its generalizability may be best understood from Vapnik's risk formula, which provides an upper bound on the total risk of a learning machine in terms of its empirical risk and structural risk for a given number of training samples. It shows that there exists an optimal VC dimension for which upper bound on total risk is minimum and if we increase machine capacity beyond this, bound on total risk exhibits increasing trend, leading to poor generalization. This problem is further aggravated for cases where training data is having class imbalance and for some deep learning architectures which are in general difficult to stabilize and control such as generative adversarial networks (GANs).

Several attempts has been made in literature to estimate bounds on structural risk, but they do not translate into a tractable problem to be optimized in terms of learning parameters of the machine. The recently proposed Minimal Complexity Machine (MCM) has shown that a bound on the VC dimension can be directly incorporated into the objective function of a learning machine, resulting in better generalizing capabilities and sparser solutions. The present

work focuses primarily on a minimal complexity implementation of Neural Networks, termed as the Low Complexity Neural Network (LCNN). This is done by deriving a differentiable and scalable loss function, that not only minimizes the empirical risk, but also the structural risk of a classifier. This implementation was extended to various deep learning architectures and tested on multiple benchmark datasets.

The Low Complexity loss function was also derived and implemented for Generative Models. It was shown that by introducing complexity term in Generative Adversarial Networks (GANs) and its variants, training stability and quality of generated image was improved and issue of mode collapse was avoided.

Next, Twin Neural Networks, a novel neural network architecture for handling class imbalance is described. Proposed class of machine not only provides a scalable framework for large sized datasets but also keeps structural risk at check resulting in good generalization. It was shown over several benchmark datasets with very high class imbalance that Twin Neural Networks produces better results than that of competing methods. A real world application of Twin Neural Network for classification of ElectroEncephaloGram (EEG) signals is also discussed.

Subsequent to this, a direct realization of the Minimal Complexity Machine (MCM) was implemented via a neurodynamical system. The proposed system solves coupled differential equations in a multilayer neural network setting, enabling stochastic learning and implicit kernel optimization. The resulting system converges to MCM solution providing better generalization and scalability. Numerical experiments on benchmark datasets show that the proposed approach is scalable and accurate, and learns models with improved accuracies and fewer support vectors.

The last part of the thesis takes an alternative route to Risk Minimization

via Data Augmentation. The proposed method is termed as Eigenbag, which presents framework for augmenting data at scale. This becomes important for scenarios, where the number of samples available for training are small. Eigenbag generates new data samples in a lower dimensional space via bagging and projects it back to the original dimension. A deep Learning version of Eigenbag is also discussed that uses Convolutional Autoencoder to increase training sample size at scale. The proposed method is also tested against competing methods.

सार

एक लर्निंग मशीन की वापनीक-चेरवोनेनकिस (वी सी) आयाम, मॉडल जटिलता नियंत्रण के क्षेत्र में एक महत्वपूर्ण अवधारणा है। वीसी आयाम एक लर्निंग मशीन की क्षमता का भी प्रतिनिधित्व करता है| प्रशिक्षण के आंकड़ों की उपलब्धता में वृद्धि और उसके बाद डीप लर्निंग आर्किटेक्चर के आगमन के साथ, एक लर्निंग मशीन की क्षमता नियंत्रण और भी महत्वपूर्ण हो गयी है। मशीन की क्षमता सीधे इसकी सामान्यता को प्रभावित करती है, जिसके कारण अनुकूलतम मान से अधिक क्षमता वाली मशीनें ओवेरफिट होने के कारण खराब परीक्षण प्रदर्शन करती हैं ।

एक मशीन की क्षमता और इसकी सामान्यता के बीच संबंध को सबसे अच्छी तरह से वापनीक के जोखिम सूत्र द्वारा समझा जा सकता है जो किसी प्रशिक्षण नमूनों की संख्या के लिए अपने अनुभवजन्य जोखिम और संरचनात्मक जोखिम के संदर्भ में एक ऊपरी सीमा प्रदान करता है। यह दर्शाता है कि एक इष्टतम वीसी आयाम मौजूद है, जिसके लिए कुल जोखिम पर ऊपरी सीमा न्यूनतम है और अगर हम इससे परे मशीन की क्षमता बढ़ाते हैं, तो कुल जोखिम की सीमा बढ़ने लगती है, जिससे सामान्यीकरण प्रभावित होता है। यह समस्या उन मामलों के लिए और बढ़ जाती है जहाँ प्रशिक्षण डेटा में कक्षा असंतुलन है और कुछ डीप लर्निंग आर्किटेक्चर के लिए जो सामान्य रूप से स्थिर और नियंत्रित करने मुश्किल हैं जैसे कि जेनरेटिव अड्वॅरसेरिअल नेटवर्क्स(GANs) |

संरचनात्मक जोखिम पर सीमा का अनुमान लगाने के लिए साहित्य में कई प्रयास किए गए हैं, लेकिन वे मशीन के सीखने के मापदंडों के संदर्भ में अनुकूलित किए जाने योग्य ट्रैक्टेबल समस्या में परिवर्तित नहीं होते हैं। हाल ही में प्रस्तावित मिनिमल कम्प्लेक्सिटी मशीन (एमसीएम) ने दिखाया है कि वीसी आयाम को सीधे एक सीखने की मशीन के उद्देश्य में शामिल किया जा सकता है जिसके परिणामस्वरूप सामान्यीकरण क्षमताओं और विरल समाधानों को बेहतर बनाया जा सकता है। वर्तमान कार्य मुख्य रूप से न्यूरल नेटवर्क के न्यूनतम जटिलता कार्यान्वयन पर केंद्रित है जिसे एक डिफ्रेंसिअबल और स्केलेबल लॉस फ़ंक्शन को प्राप्त करके लो कम्प्लेक्सिटी न्यूरल नेटवर्क (LCNN) कहा गया है, जो न केवल एक क्लासिफायरियर के अनुभवजन्य जोखिम को कम करता है बल्कि संरचनात्मक जोखिम को भी कम करता है। इस कार्यान्वयन को विभिन्न डीप लर्निंग आर्किटेक्चर में कार्यान्वित गया था और बेंचमार्क डेटासेट पर परीक्षण किया गया। लो कॉम्प्लेक्सिटी लॉस फंक्शन जेनरेटिव मॉडल्स के लिए भी कार्यान्वित किया गया । यह प्रस्तुत किया गया था कि जनरेटिव अड्वॅरसेरिअल नेटवर्क्स (GANs) और इसके वेरिएंट्स में लो कम्प्लेक्सिटी का समावेश करने से, प्रशिक्षण की स्थिरता और उत्पन्न छवि की गुणवत्ता में सुधार आया और मोड कोलैप्स के मुद्दे से भी बचा गया।

तदोपरांत, वर्ग असंतुलन से निपटने के लिए , द्विन न्यूरल नेटवर्क्स, एक नवीन न्यूरल नेटवर्क का वर्णन किया गया है। मशीन का प्रस्तावित वर्ग न केवल बड़े आकार के डेटासेट के लिए एक स्केलेबल फ्रेमवर्क प्रदान करता है, बल्कि संरचनात्मक रूप से जोखिम भी कम रखता है जिसके परिणामस्वरूप अच्छा सामान्यीकरण होता है। यह उच्च वर्ग असंतुलन वाले कई बेंचमार्क डेटासेट पर दिखाया गया था कि द्विन न्यूरल नेटवर्क प्रतिस्पर्धी तरीकों की तुलना में बेहतर परिणाम देता है। इलेक्ट्रोएन्सेफलोग्राम (EEG) संकेतों के वर्गीकरण के लिए द्विन न्यूरल नेटवर्क का एक वास्तविक अनुप्रयोग भी प्रदर्शित किया गया है।

इसके बाद, न्यूरोडायनामिकल प्रणाली के माध्यम से मिनिमल कम्प्लेक्सिटी मशीन (एमसीएम) को लागू किया गया है । प्रस्तावित प्रणाली एक बहुपरत न्यूरल नेटवर्क सेटिंग में युग्मित समीकरणों को हल करती है और स्टोकास्टिक सीखने और अंतर्निहित कर्नेल अनुकूलन को सक्षम करती है। परिणामस्वरूप प्रणाली एम सी एम समाधान की ओर अभिसरित होकर बेहतर सामान्यीकरण और स्केलेबिलिटी प्रदान करती है। बेंचमार्क डेटासेट पर संख्यात्मक प्रयोग बताते हैं कि प्रस्तावित दृष्टिकोण स्केलेबल और सटीक है, और बेहतर सटीकता और कम सपोर्ट वेक्टर्स के साथ मॉडल सीखता है।

थीसिस का अंतिम भाग डेटा ऑगमेंटेशन तकनीक के माध्यम से जोखिम न्यूनतमकरण के लिए एक वैकल्पिक मार्ग लेता है। प्रस्तावित विधि को आएगॉनबैग कहा गया है, जो बड़े पैमाने पर डेटा को बढ़ाने के लिए रूपरेखा प्रस्तुत करता है। यह उन परिदृश्यों के लिए और भी महत्वपूर्ण हो जाता है, जहां प्रशिक्षण के लिए उपलब्ध नमूनों की संख्या बहुत कम है। आएगॉनबैग एक कम आयामी स्थान में बैगिंग के माध्यम से नए डेटा नमूने उत्पन्न करता है और इसे मूल आयाम पर प्रक्षेपित करता है। आएगॉनबैग

का एक डीप लर्निंग संस्करण भी प्रस्तुत किया गया है , जो बड़े पैमाने पर प्रशिक्षण के नमूने को बढ़ाने के लिए कोंवोलूशनल ऑटोकेनडर का उपयोग करता है। प्रतिस्पर्धी तरीकों के खिलाफ प्रस्तावित विधि का परीक्षण भी किया गया है।

Contents

List of Figures

List of Tables