Preventing Model Overfitting and Underfitting in Convolutional Neural Networks

Article *in* International Journal of Software Science and Computational Intelligence · October 2018

DOI: 10.4018/IJSSCI.2018100102

Andrei Dmitri Gavrilov
University of British Columbia - Vancouver

Jack Deng
China Jiliang University

International Journal of Software Science and Computational Intelligence
Volume 10 • Issue 4 • October-December 2018

Preventing Model Overfitting and Underfitting in Convolutional Neural Networks

Andrei Dmitri Gavrilov, University of British Columbia, Calgary, Canada

Alex Jordache, University of British Columbia, Calgary, Canada

Maya Vasdani, University of British Columbia, Calgary, Canada

Jack Deng, University of British Columbia, Calgary, Canada

ABSTRACT

The current discourse in the machine learning domain converges to the agreement that machine learning methods emerged as some of the most prominent learning and classification approaches over the past decade. The CNN became one of most actively researched and broadly-applied deep machine learning methods. However, the training set has a large influence on the accuracy of a network and it is paramount to create an architecture that supports its maximum training and recognition performance. The problem considered in this article is how to prevent overfitting and underfitting. The deficiencies are addressed by comparing the statistics of CNN image recognition algorithms to the Ising model. Using a two-dimensional square-lattice array, the impact that the learning rate and regularization rate parameters have on the adaptability of CNNs for image classification are evaluated. The obtained results contribute to a better theoretical understanding of a CNN and provide concrete guidance on preventing model overfitting and underfitting when a CNN is applied for image recognition tasks.

KEYWORDS

Cognitive Systems, Convolutional Neural Networks, Image Processing, Ising Model, Learning Rate, Machine Learning, Overfitting, Regularization Rate, Underfitting

1. INTRODUCTION

Over the past decade, machine learning evolved into an increasingly powerful approach among other artificial intelligence methods, where computers can discover patterns and learn from the collected data to make intelligent autonomous decisions. A CNN (Convolution Neural Network) is a class of deep learning neural networks that can be applied to various classification and recognition tasks, such as identifying a human in a surveillance video, recommending specific products for consumers, observing interesting weather phenomena, or identifying key chemical structure in various drug developments (Wang et al., 2017).

DOI: 10.4018/IJSSCI.2018100102

However, the CNN approach performance depends on both the data set composition and the parameter set. Data set training has a large influence on the accuracy of a network, and hence it is paramount to create a network architecture that prevents overfitting and underfitting (Faussett, 2004). By randomly trimming the data during training and providing more robust data sets, the network can become less reliant on similar pieces of data. Hence, this would improve its overall capability for higher accuracy image recognition.

This paper investigates the problem of overfitting and underfitting, analyzing performance of CNN based image recognition algorithms using insights from the Ising model. We propose to use the analogy to the square-lattice array, in order to determine how the statistical mechanics of the Ising model phase transition are used in tuning parameters of the CNN. Once a training set of such data is complete, we explore the convolutional layers that propagate the neuron values, similar to the propagation of spins. The experimental results conducted on CIFAR image database demonstrate that a low regularization rate and learning rate yields overfit data, high regularization rate yields partially fit data, and high learning rate yields unfit data. Thus, this research provides insights that can assist in preventing model overfitting and underfitting when CNN is utilized in various image recognition applications.

This article is an extended version of our previous research which was published at the 2018 International Conference on Cognitive Sciences & Cognitive Computing 2018 and was presented at Berkeley, California, in July 2018 (Gavrilov et al., 2018).

2. BACKGROUND RESEARCH

Cognitive computing and cognitive architectures recently emerged as powerful tools to tackle complex large-scale real-life problems in the presence of uncertainty and variable data quality (Tian et al. 2012), (Wang et al., 2013), (Wang et al., 2016). Popular approaches that assist in building cognitive models, which can simulate human thought process, include deep machine learning methods, artificial neural networks (ANN), convolution neural networks (CNN), neuro-linguistic programming (NLP) and sentiment analysis. They have been successfully applied to various intelligent systems in the fields of computer graphics, robotics, knowledge representation, virtual reality, situation awareness, decision-support systems, medicine and many other areas (Wang et.al., 2017), (Gavrilova et al., 2017), (Montero-Obasso et al., 2012). One of the fastest growing domains where notable progress has been made using cognitive, fuzzy and multi-modal architectures is biometric security and image processing (Browne & Ghidary, 2003), (Han & Bhanu, 2006), (Monwar et al., 2011), (Yuan et al., 2008). Over past couple of years, there has been a significant surge in adapting machine-learning methods for image recognition. The introduction of CNN created excitement in image processing research community, with new opportunities to significantly increase image identification rate with a fraction of computational resources, thus making the recognition process more accurate and less resource demanding.

A CNN (Convolutional Neural Network) is one of actively researched and broadly applied deep machine learning methods. A CNN is composed of a feed-forward neural network that takes in images as inputs, and outputs a probability value associated to a class that best describes the image. As well, it is constructed of multiple layers, which include convolutional, max-pooling and fully connected layers, that perform classification tasks. A CNN learns a set of features important for image recognition that previously were hand-picked by traditional sequential algorithms (Krizhevsky et al., 2017). There have been numerous examples of CNNs performing very well for a variety of applied problems (Wang & Tian, 2013). For example, they have been successfully used in engineering domains for fuzzy-logic control and decision systems (Lin & Lee, 1991). The very recent overview of their applications and open questions can be found in 2017 research article by (McCann et al. 2017). In this state-of-the-art review, authors describe a gamut of uses of CNNs to solve both direct and inverse problems in imaging. Authors also point out that once it became possible to train deep