A survey of deep neural network architectures and their applications

Weibo Liu\textsuperscript{a}, Zidong Wang\textsuperscript{a,}\textsuperscript{x}, Xiaohui Liu\textsuperscript{a}, Nanyin Zeng\textsuperscript{b}, Yurong Liu\textsuperscript{c,d}, Fuad E. Alsaadi\textsuperscript{d}

\textsuperscript{a} Department of Computer Science, Brunel University London, Uxbridge, Middlesex, UB8 3PH, United Kingdom
\textsuperscript{b} Department of Instrumental and Electrical Engineering, Xiamen University, Xiamen 361005, Fujian, China
\textsuperscript{c} Department of Mathematics, Yangzhou University, Yangzhou 225002, China
\textsuperscript{d} Communication Systems and Networks (CSN) Research Group, Faculty of Engineering, King Abdulaziz University, Jeddah 21589, Saudi Arabia

\textbf{A B S T R A C T}

Since the proposal of a fast learning algorithm for deep belief networks in 2006, the deep learning techniques have drawn ever-increasing research interests because of their inherent capability of overcoming the drawback of traditional algorithms dependent on hand-designed features. Deep learning approaches have also been found to be suitable for big data analysis with successful applications to computer vision, pattern recognition, speech recognition, natural language processing, and recommendation systems. In this paper, we discuss some widely-used deep learning architectures and their practical applications. An up-to-date overview is provided on four deep learning architectures, namely, autoencoder, convolutional neural network, deep belief network, and restricted Boltzmann machine. Different types of deep neural networks are surveyed and recent progresses are summarized. Applications of deep learning techniques on some selected areas (speech recognition, pattern recognition and computer vision) are highlighted. A list of future research topics are finally given with clear justifications.

1. Introduction

Machine learning techniques have been widely applied in a variety of areas such as pattern recognition, natural language processing and computational learning. With machine learning techniques, computers are endowed with the capability of acting without being explicitly programmed, constructing algorithms that can learn from data, and making data-driven decisions or predictions. During the past decades, machine learning has brought enormous influence on our daily life with examples including efficient web search, self-driving systems, computer vision, and optical character recognition. In addition, by adopting machine learning methods, the human-level artificial intelligence (AI) has been improved as well, see [101,137,165] for more discussions. Nevertheless, when it comes to the human information processing mechanisms (e.g. speech and vision), the performance of traditional machine learning techniques are far from satisfactory. Inspired by deep hierarchical structures of human speech perception and production systems, the concept of deep learning algorithms was introduced in the late 20th century. Breakthroughs on deep learning have been achieved since 2006 when Hinton proposed a novel deep structured learning architecture called deep belief network (DBN) [59]. The past decade has seen rapid developments of deep learning techniques with significant impacts on signal and information processing. Research on neuromorphic systems also supports the development of deep network models [75]. In contrast to traditional machine learning and artificial intelligence approaches, the deep learning technologies have recently been progressing massively with successful applications to speech recognition, natural language processing (NLP), information retrieval, compute vision, and image analysis [91,125,159].

The concept of deep learning originated from the study on artificial neural networks (ANNs) [60]. ANNs have become an active research area during the past few decades [63,162,166,167,175]. To construct a standard neural network (NN), it is essential to utilize neurons to produce real-valued activations and, by adjusting the weights, the NNs behave as expected. However, depending on the problems, the process of training a NN may take long causal chains of computational stages. Backpropagation is an efficient gradient descent algorithm which has played an important role in NNs since 1980. It trains the ANNs with a teacher-based supervised learning approach. Although the training accuracy is high, the performance of backpropagation when applied to the testing data might not be satisfactory. As backpropagation is based on local gradient information with a random initial point, the algorithm often gets trapped in local optima. Furthermore, if the size of the training data is not big enough, NNs will face the problem of over-
fitting. Consequently, other effective machine learning algorithms such as support vector machine (SVM), boosting and K-nearest neighbour (KNN) have been adopted to obtain global optimum with lower power consumption. In 2006, Hinton [59] proposed a new training method (called layer-wise-greedy-learning) which marked the birth of deep learning techniques. The basic idea of the layer-wise-greedy-learning is that unsupervised learning should be performed for network pre-training before the subsequent layer-by-layer training. By extracting features from the inputs, the data dimension is reduced and a compact representation is hence obtained. Then, exporting the features to the next layer, all of the samples will be labeled and the network will be fine-tuned with the labeled data. The reason for the popularity of deep learning is twofold: on one hand, the development of big data analysis techniques indicates that the overfitting problem in training data can be partially solved; on the other hand, the pre-training procedure before unsupervised learning will assign non-random initial values to the network. Therefore, a better local minimum can be reached after the training process and a faster converge rate can be achieved.

Up to now, the research on deep learning techniques has stirred a great deal of attention and a series of exciting results have been reported in the literatures. Since 2009, the ImageNet’s competition has attracted a great many computer vision research groups throughout the world from both academia and industry. In 2012, the research group led by Hinton won the competition of ImageNet Image Classification by using deep learning approaches [86]. Hinton’s group attended the competition for the first time and their results were 10% better than that in the second place. Both Google and Baidu have updated their image search engines based on Hinton’s deep learning architecture with great improvements in searching accuracy. Baidu also set up the Institute of Deep Learning (IDL) in 2013 and invited Andrew Ng, the associate professor at Stanford University, as the Chief Scientist. In March 2016, a Go Game match was held in South Korea by Google’s deep learning project (called DeepMind) between their AI player AlphaGo and one of the world’s strongest players Lee Se-dol [140]. It turned out that AlphaGo, adopting deep learning techniques, showed surprising strength and beat Lee Se-dol by 4:1. In addition, deep learning algorithms have also shown outstanding performance in predicting the activity of potential drug molecules and the effects of mutations in non-coding DNA on gene expression.

With rapid development of computation techniques, a powerful framework has been provided by ANNs with deep architectures for supervised learning. Generally speaking, the deep learning algorithm consists of a hierarchical architecture with many layers each of which constitutes a non-linear information processing unit. In this paper, we only discuss deep architectures in NNs. Deep neural networks (DNNs), which employ deep architectures in NNs, can represent functions with higher complexity if the numbers of layers and units in a single layer are increased. Given enough labeled training datasets and suitable models, deep learning approaches can help humans establish mapping functions for operation convenience. In this paper, four main deep architectures are recalled and other methods (e.g. sparse coding) are also briefly discussed. Additionally, some recent advances in the field of deep learning are described.

The purpose of this article is to provide a timely review and introduction on the deep learning technologies and their applications. It is aimed to provide the readers with a background on different deep learning architectures and also the latest development as well as achievements in this area. The rest of the paper is organized as follows. In Sections 2–5, four main deep learning architectures, which are restricted Boltzmann machines (RBMs), deep belief networks (DBNs), autoencoder (AE), and convolutional neural networks (CNNs), are reviewed, respectively. Comparisons are made among these deep architectures and recent developments on these algorithms are discussed. The applications of those deep architectures are highlighted in Section 6. Conclusions and future topics of research are presented in Section 7.

2. Deep learning architectures: restricted Boltzmann machine

2.1. The motivation

In this part, a brief review of RBMs is given. RBMs are widely used in deep learning networks on account of their historical importance and relative simplicity. The RBM was first proposed as a concept by Smolensky, and has become prominent since Hinton published his work [59] in 2006. RBMs have been used to generate stochastic models of ANNs which can learn the probability distribution with respect to their inputs. RBMs consist of a variant of Boltzmann machines (BMs). BMs can be interpreted as NNs with stochastic processing units connected bidirectionally. Since it is difficult to learn aspects of an unknown probability distribution, RBMs have been proposed to simplify the topology of the network and to enhance the efficiency of the model. It is well recognized that an RBM is a special type of Markov random fields with stochastic visible units in one layer and stochastic observable units in the other layer.

2.2. The structure and the algorithm

As shown in Fig. 1, the neurons are restricted to form a bipartite graph in an RBM. It can be seen that there is a full connection between the visible units and the hidden ones, while no connection exists between units from the same layer [165]. To train an RBM, the Gibbs sampler is adopted. Starting with a random state in one layer and performing Gibbs sampling, we can generate data from an RBM. Once the states of the units in one layer are given, all the units in the other layers will be updated. This update process will carry on until the equilibrium distribution is reached. Next, the weights within an RBM are obtained by maximizing the likelihood of this RBM. Specifically, taking the gradient of the log-probability of the training data, the weights can be updated according to:

$$\frac{\partial \log p(v^0)}{\partial \omega_{ij}} = \langle v_i^0 h_j^0 \rangle - \langle v_i^m h_j^m \rangle,$$

(1)

where $\omega_{ij}$ represents the weight between the visible unit $i$ and the hidden unit $j$, $(v_i^0 h_j^0)$ and $(v_i^m h_j^m)$ are the correlations when the visible and hidden units are in the lowest layer and the highest layer,
دریافت فوری
متن کامل مقاله

امکان دانلود نسخه تمام متن مقالات انگلیسی
امکان دانلود نسخه ترجمه شده مقالات
پذیرش سفارش ترجمه تخصصی
امکان جستجو در آرشیو جامعی از صدها موضوع و هزاران مقاله
امکان دانلود رایگان ۲ صفحه اول هر مقاله
امکان پرداخت اینترنتی با کلیه کارت های عضو شتاب
دانلود فوری مقاله پس از پرداخت آنلاین
پشتیبانی کامل خرید با بهره مندی از سیستم هوشمند رهگیری سفارشات