CREATION, ANALYSIS AND VERIFICATION OF DEEP LEARNING MODELS USING MULTI-VARYING DATASETS

KAUSHIK.G.VISWANATH
(RA1611008020089)
IT-'B' IIIrd YEAR
SRM INSTITUTE OF SCIENCE AND TECHNOLOGY
RAMAPURAM, CHENNAI-600 089

ARCHISMAN SARKAR
(RA1611008020118)
IT-'B' IIIrd YEAR
SRM INSTITUTE OF SCIENCE AND TECHNOLOGY
RAMAPURAM, CHENNAI-600 089

BRYAN JOHN SAMUEL.I
(RA1611008020090)

MS.ABINAYAA
MENTOR

IT-'B' IIIrd YEAR
SRM INSTITUTE OF SCIENCE AND TECHNOLOGY
RAMAPURAM, CHENNAI-600 089

ABSTRACT

The idea behind Deep Learning is to not only to provide automation but to also make humans concur to the fact that the field of Deep Learning is also to reduce human effort and errors to a bare minimum. Deep Learning has become one of the most in-demand fields in modern Information Technology and Computing principles. It also forms the cornerstone of this project undertaken by us. The prospect of working with something new and exciting has provided an opportunity of learning and purpose for us. The desire is to create systems that can provide recursive solutions based on user choices that form part of many modern datasets upon which real-time solutions can be chalked up using enhanced and specialised Deep Learning Algorithms the usual ones being Linear Regression and Logistical Regression. We will create recursive use cases based on these algorithms. We intend to download the datasets and the programs from GitHub Repositories and work on them to create meaningful data models and through their analysis and verification (using manual validation techniques). The constraint being Deep Learning is still very much in its nascent and research stages and this is also more or less a research concept, so our focus still remains in learning about this field through experimental procedures.

I. INTRODUCTION

The objective is to create specialized deep learning models that can readily observe, understand, evaluate and generate meaningful trends based on modern researches from various fields. We come across many situations in today's world where existing trends are analysed and ideas for promotion are presented. From analysing product sales for a

supermarket chain to analysing social media sentiments we collaborate to put up these trends in a system that can analyse them to provide meaningful situations for upcoming times or to provide respite should the situations repeat themselves. Deep Learning is part of a broader family of machine learning methods based on learning data representations, as opposed to task-specific algorithms. Learning can be supervised, semi-supervised or unsupervised. Deep Learning forms the analytical part of AI concerned with creating and evaluating neural network models by emulating patterns with the given data. Being in a very nascent stage, Deep Learning forms the subset upon which ensuing fields like Machine Learning and AI exist. The ability to create procreative and recursive models upon repeated training of data to match the result for a particular purpose sees the scope of Deep Learning soar to newer heights. The applications of Deep Learning are wide-ranged and multi-disciplinary the reason for which there is high level of research and demand going on. From market trends research to providing user preference based advertisements from analysing repetitive user search trends on the internet to provide marketing feedback. Such analysis can help marketing companies understanding customer sentiment and thereby base their marketing on the same lines as deemed fit by the customer. Also for FMCG companies, it can be the understanding what standard of products a consumer has special need for or has liking to. The inclination is towards assessing what a particular situation of any nature demands and the progressive solutions that can be generated for any purpose so deemed fit by the particular individual for whom the outputs are solicited.

MATERIALS AND METHODS

This project initially demands proactive knowledge of statistics and understanding of the science of probability. Also required is the ability to analyse and understand the repetitive trends that can be generated using the data. Most important of all is to root out the basic requirement for this initiative: data!! Loads of data are paramount for this project and for learning purposes we will utilize various arbitrary datasets available on the internet through various online data repositories and so forth. The project demands cohesive partnership between Data Science and Deep Learning. Preparation of data is going to be in the format of testing and model data. Part of the data is to be allocated to testing any possible outcomes deemed feasible for the developer or for a particular user. The models are going to be in the form of Python programs coded to assess and analyse datasets and create and fit a model that appeases any output desired for the particular dataset. Graphical representation of the solution is designed as the ultimate output although simpler programmable outputs can also be used but demand lesser desirability of choice. The existing systems using Deep Learning models for assessing various datasets achieve a typical accuracy of 66-71% at a modest scale.

MODULES AND IMPLEMENTATION

We require specialized mathematical and analytical library files for execution of deep learning models, like Theano (for backend usage), Tensorflow and Keras (for mathematical and evaluation purposes). For the graphical representation of the output results following dataset evaluation, we use TensorBoard module provided within the Tensorflow backup files itself. The idea behind using Tensorflow ahead of Theano is the ability of Tensorflow to provide additional functionality for development processes as well as they have wider appeal when it comes to creating production systems using Deep Learning models. Tensorflow also uses a higher level of NumPy and SciPy library files. Also to mention is the use of Keras Python Library that provides a much more convenient way to create recursive deep learning models. Keras can act as the frontend idea behind creation of the models with Theano and/or Tensorflow acting as the backend much like a neutralizing interface. A specific function of the Numpy library file is the presence or rather the provision of the "Matplotlib" function to represent the physical data or outputs obtained in a graphical (or) diagrammatical manner. The Matplotlib function is used as an output function for which calling the Numpy library is imported specifically much similar to calling #include<conio.h> for calling the getch() function in C/C++ programming.

A typical Deep Learning model requires data, lots of data. Datasets are typically raw collections of data which have the capability to provide conclusive results for evaluation owing to which multiple trends can be drawn out worth the need for the particular individual. The datasets form the crux of this project, and following the principle that the scope of this project is still in research for that purpose, the datasets will be obtained via online repositories or storehouses of such information available all over the internet. We will be individually calling these datasets using their extension calls (.csv files).

The life cycle of the creation of a DL model requires the following steps:-

*) Importing all libraries

*) Loading the datasets

*) Define Model

*) Compile Model

*) Fit Model

*) Evaluate Model

*) Binding of all the above steps

The above steps are to be followed in sequential manner and are to be followed to the letter. Without delving too much into specifics, we can assure that following these steps will lead us to creating proper DL models. The representative measures require separate set of steps for showcasing purposes. As mentioned above we will use Tensorboard modules for the graphical purposes.

RESULTS

- The proposed system is planned to be designed in a manner of achieving a minimum accuracy of at least 78% in extracting and parsing of data and creation of recursive models to denote solutions for the same.
- As proposed above, graphical representation is deemed as the desirable form of output.
- Our major source of inputs comes in the form of datasets and meaningful charts are obtained for the same as output.
- Usage of Supervised learning techniques denotes a high level of data structuring using labeled data.
- Usage of unsupervised learning techniques will help us serialize non-linear and non-sequential data in a structured manner much similar to the levels of supervised learning.

REFERENCES

[1] G.Zaccone, Md.R.Karim, A.Menshawy, “Deep Learning with Tensorflow”, April 2017.

[2] J.Brownlee, “Deep Learning with Python: Develop Deep Learning Models using Theano and Tensorflow using Keras” v1.6 2016.

[3] A.Balaji, T.Ramanathan, V.Sonathi,”Chart-Text: A Fully Automated Chart Image Descriptor” arXiv: 1812.10636 December 2018.

[4] Margaret M.R, “Anaconda, Jupyter Notebook, Tensorflow and Keras for Deep Learning”

[5] For online dataset repositories: www.github.com

[6] Basic concepts of Deep Learning learnt from: www.DeZyre.com