

Introduction

Gradient Based Learning Applied to Document Recognition, by Yann LeCun, Leon Bottou, Yoshua Bengio and Patrick Haffner, is a review of the various technologies and techniques implemented in document recognition systems, comparing their relative efficiencies and determining which are the best in terms of accuracy, computation time or memory usage.

It is a high quality evaluation of the merits and flaws of Convolutional Neural Networks and Graph Transformer Networks when applied to real-world applications, and the experiments, results, and conclusions are clearly presented. The paper is only let down by its age, as it obviously fails to reflect advances made in the two decades since publication.

Content

This paper contains an overview of the topic of Convolutional Neural Networks, including within its scope their construction, functionality, and efficiency as a machine learning mechanisms, especially in comparison to more traditional methods of machine learning. Specifically, it examines the problem of character recognition, a problem historically too varied to allow for a completely handwritten solution; that is, a problem particularly well suited to the implementation of machine learning mechanisms.

The document discusses the advantages provided by the Convolutional Neural Network's ability to generate its own features, and describes a character recognition case study used as evidence to prove that feature extraction performed entirely by machine learning can be superior to hand-crafted feature extraction methods, achieving a greater level of accuracy in predictions. In addition, the paper notes that the ability of a machine learning system to generate its own features allows it to, after some iteration, perform on unseen problems with high accuracy, which is generally not possible for hand-crafted feature extraction methods of any particular specificity.

The document further describes the general design paradigms of machine learning systems and how Convolutional Neural Networks provide advantages in this area as well, such as how the two step design of recognition systems can be replaced through the architecture of Convolutional Neural Networks. The document utilizes a case study demonstrating a wide variety of Convolutional Neural Network architectures in addition to more traditional methods to demonstrate the differences in accuracy, memory efficiency, and runtime.

The second half of the document details the concept of a Graph Transformer Networks, multi-module systems designed to minimize error across the entire network. It describes a case where this technique is useful, which is stepping from individual character recognition to recognition of whole words and sentences, and details other possible uses as well as evolutions upon the idea.

It closes by bringing all these concepts together in order to solve the problem first posed in its title, which is document recognition, with a case study on a real application, which is cheque recognition in the banking industry, and an analysis of how the implementation of Convolutional Neural Networks and Graph Transformer Networks has assisted in achieving this goal.

Innovation

At the time of its publication, this research paper was one of the first papers to provide any in-depth analysis on the construction or functionality of Convolutional Neural Networks and Graph Transformer Networks in comparison to more traditional methods, using common case studies as evidence.

The first half of the paper puts forth the utility of Convolutional Neural Networks eliminating the need for hand-crafted feature extractors, which has significant advantages for prediction accuracy in both seen and unseen problems. Whilst this is a well reasoned argument, clearly presented and supported by comprehensive comparative analysis, the paper places little weight on the downsides of Convolutional Neural Networks compared to more traditional methods.

Though it is noted and shown in diagrams that Convolutional Neural Networks require much greater computational and memory capacity, this is largely waved away with the remark that as computers develop these issues will become less significant. Whilst this is certainly true, at the time the paper was published the massively increased requirements to run a Convolutional Neural Network would have been prohibitive, compared to any other kind of classifier, and this should have warranted some mention.

The second half introduces Graph Transformer Networks and asserts its success in scenarios where the system's functional architecture is dynamically altered by new input, such as document analysis. Significant effort is invested into explaining and demonstrating the functionality of this system, and proving its effectiveness using real world case examples. Again, in addition to presently achievable uses, the paper highlights elements of the system that could provide a basis for future research or development.

However, once again only a little attention was given to the problems and limitations of the system. This technique requires a vast amount of training data, which itself generally needs to be prepared to a higher level than usual training data sets. While not an inherent design flaw with the system itself, this makes implementing it into unseen problems, and even many commonly known problems, a difficult process. Overcoming this issue would require significant investment of time and effort to adequately prepare the training data for use.

Technical Quality

The experiments performed are clearly documented and explained, ensuring they could be replicated by a sufficiently motivated reader with minimal additional effort beyond following the instructions in the paper. The results, especially the more interesting elements, are explained in detail, and all conclusions drawn are well reasoned and supported with evidence.

For the first part of the paper, focusing on the utility of Convolutional Neural Networks, twenty seven different classification methods were used for comparison purposes, ensuring that there would be a range of test results and reducing the risk of fluke results severely impacting the overall comparison.

All classifiers were trained and tested using the same database, ensuring that there is no bias induced by differences in the training data. The first diagram shows the raw error rate of the different classifiers, but additional diagrams are used to further examine other differences, especially those that do not apply to all classifiers (such as the diagram displaying rejection performance, the number of test patterns that were rejected as they could not be accurately guessed).

In addition to the accuracy ratings, some attention is paid to the relative computational and storage requirements of different classifiers, although this is only given token attention in the text.

The second half of the paper, focused on multi-module systems and Graph Transformer Networks, is also well constructed, with a good textual explanation of the structure and concepts, supplemented with diagrams and pseudocode built around a document interpreter used as a training/testing problem.

The instructions assume a significant knowledge of implementing machine learning but the particular details of the Graph Transformer Network are clearly provided for audiences to attempt construction of their own. Whilst this does limit the replicability of this part of the experiment, it is clear that significant effort was expended by the authors in an attempt to make it as easy as practical for the experiment to be replicated.

Application and X-Factor

Though now the paper is twenty years old and far behind modern developments, it still contains a functional and detailed overview of the basic operation and structure of Convolutional Neural Networks. It provides an excellent starting point for people new to the field and invites significant future development.

The paper makes several notes of industries where Convolutional Neural Networks and Graph Transformer Networks are already in use, explaining and demonstrating the efficiency of a variety of Convolutional Neural Networks over hand-crafted machine learning systems using real-world case studies. Because of the architecture of Convolutional Neural Networks there are a vast number of potential variations upon existing architectures, and an endless series of new architectures of varying complexities and sizes, opening the path for a significant amount of further research and investigation in the development and application of new architectures.

Further, the paper does make note that there are other applications that could benefit from these techniques, allowing the potential for significant future development built off this technique. No specific examples are provided, but the potential applications of automatic learning are significant across many fields. Broadly speaking, any situation where a problem's domain knowledge or state information can be reduced to or rephrased as a graph stands to benefit from the application of this technique.

Presentation

The paper is well written with no exceptionally obvious spelling or grammar errors. The sentence structure is clear throughout the document, and the flow from one section to the next is natural and understandable, with no significant leaps in continuity. The arguments are clearly laid out, and proofs and evidence are linked back to the point they prove.

All diagrams inserted into the text are well structured, neatly drawn, and labelled, and their purpose and content is clearly explained both in the text and in the captions. The diagrams are kept simple, preventing them from becoming too noisy or difficult to read to be useful.

Where applied, mathematical formulas are written neatly, and the more complicated ones are explained and labelled appropriately. The language is formal but not too dense or unwieldy to understand.

References

Goodfellow, I., Bengio, Y., and Courville, A. (2016) *Deep Learning*. MIT Press.

Sandler, M., Howard, A.G., Zhu, M., Zhmoginov, A., and Chen, L. (2018) Inverted residuals and linear bottlenecks: Mobile networks for classification, detection and segmentation. *Computing Research Repository*

Zeiler, M.D. and Fergus, R. (2013) Visualizing and Understanding Convolutional Networks. *arXiv e-prints*

Zhou, H., Alvarez, J.M., and Porikli, F.M. (2016) Less is more: Towards compact CNNs. *European Conference on Computer Vision*.

Self-Links

PDF

https://github.com/C-Corby-016/UTS_ML2019_ID12958144/blob/master/12958144_ADA_Assignment_1.pdf

IPYNB

https://github.com/C-Corby-016/UTS_ML2019_ID12958144/blob/master/12958144_ADA_Assignment_1.ipynb