Term Paper

For Research Directions

University of Nebraska at Omaha

by

Chad Crowe

November 2018

Supervisor

Dr. Hall

Acknowledgements

– To Dr. Hall and her patience and guidance in my research

Contents

List of Figures

List of Tables

0.1 Overview of Applied Machine Learning

Applied machine learning is a large topic. It takes the topics and tools of machine learning and theorizes applications for the real-world. The research explores new ways to represent, manipulate, and guide data analysis with machine learning tools. The research often creates new architectures that perform better in an applied scenario. There are many tools within machine learning. Machine learning architecture is also very flexible. For example, each machine learning model can consist of a varing number of layers, nodes, and even submodels. Moreover, machine learning behavior depends on its input. Some of the papers covered in the related work will explore how data manipulations can further the performance of existing machine learning models. For example, it has been shown that providing regularization produces a model that performs better on new data. Furthermore, there are many ways and techniques to apply regularization. Newer theories have proposed new techniques for regularization. Moreover, many of these new techniques may apply to certain types of machine learning architectures or nodes, such as new forms of regularization for LSTM nodes. Of particular interest to my research has included techniques for handling multi-varied types of input. For example, my masters thesis explored the combination of text and image data within a single machine learning model. These types of advanced machine learning pipelines drive big data, big analysis, and ways to analyze complicated data. I plan on continuing to drive my research within the direction of exploring new ways to handle various inputs with different data types. The field of applied machine learning is vast and full of opportunity for new researchers to discover new theories and make innovations in the use, application, and understanding of machine learning.

From the related work covered in this paper, there are major threads under development in the arena of applied machine learning. These pertain to machine learning with each type of data. The most popular types of data that receive attention when researchers theorize about their use and possible applications are text, audio, visual, and time-series data. These data types are popular because they are commonly available. Such data are available in public repositories of data or through application APIs, such as text and image data that is available from utilizing the Twitter API. Each data type has multiple facets. More interestingly, certain machine learning algorithms perform better and differently for particular data types. Neural Networks (NNs) perform very well on simple text learning tasks, yet struggle to learn image data. Convolutional Neural Networks (CNNs) may underperform with text-analysis, but have excelled at image-analysis. Video analysis tends to use Long-Term Short Memory nodes (LSTMs), but often incorporate CNNs, since videos are composed of many images.

While there are many techniques for analyzing different types of data, society and academia has seen an increase in the amount of available data. Moreover, not all data is the same. There are many types of video data. There is still video data, moving video data, video at night, day, in different places, and with more people. These data differences exist in text data too. A researcher can analyze text data from Twitter, which looks different than text data from a research paper, which also looks different from text data translated from audio data. The different data types and variations within data types are a challenge for researchers. The creation of theories about Twitter text data represents advances in our knowledge of social media. Creating models that are better at understanding text data that is created from audible speech can further research's understanding of text sentence structure. These examples only illustrate that there are many ways to approach each data type, and each data type provides many ways to explore and understand its data.

There are many benchmarks within the field of applied machine learning. Many of these benchmarks are the performance of well-tuned machine learning algorithms on a given set of data. It is common for papers within the field of applied machine learning to attempt to improve upon machine learning model performance benchmarks with new tools or techniques. Another common occurrence is to create an architecture that performs well on new data, or with a new aspect of data. The paper will then apply the machine learning architecture against existing benchmarks to show it performs similarly well on the data. It is common to apply the machine learning architecture to existing benchmarks as a litmus test of the machine learning model's overall performance. There are many benchmarks from repositories like Kaggle or libraries like sklearn. New research is always creating new data and creating new benchmarks.

It is worth reviewing the popular tools and approaches within the field of applied machine learning. Machine learning models take inputs. These inputs are transformed by models into outputs. There are many manipulations on the data while it passes through the models. The outputs depend on the type of machine learning. Two common tasks within machine learning are classification and regression. Classification models tend to output a probability for each classification through the final layer, known as the softmax layer. How the input is manipulated depends on the type of machine learning model. One common type of machine learning model is an Artificial Neural Network (ANNs), which apply simple functions to the inputs, such as a signoid, relu, or tanh function. The model is generally composed of a few layers, where each node is connected to each node in its subsequent layer. The network's input weights, for each layer, are adjusted based on the model's error. The Convolutional Neural Network (CNNs) is similar. The CNN is also composed of nodes. Half of the CNN learning is the same as ANNs, i.e. nodes connect to one another and go through a simple learning function like relu. The difference with CNNs is how the layers are connected. Instead of connecting all layers together, CNNs split the data up into regional chunks, e.g. the upper right array of data. The region of datas are connected. One intersting behavior from the CNN is that region sizes are halved and joined. In this way, regional behavior is captured and joined together in the output. Another major type of machine learning model is a Recurrent Neural Network (RNNs). RNNs are very similar to ANNs, except their layers are connected differently. The RNNs layers do not always connect to the next layer, i.e. the layer's output may serve as input to a previous/the same layer. This provides a cycle of input output information within the neural network. This allows each layer to have some input from future layers, which creates a sense of network memory, where previous data serves as input to the network's current behavior.

0.2 Related Work

Shen et. al explores attention mechanisms for machine learning. The subject of attention mechanisms is not well known within machine learning. It has recently attracted a large amount of attention, due it its performance and speed of computation. Since attention mechanisms are more lightweight, they train faster. The mechanism, as will be explained, still relies on nodes, and therefore has much of the flexibility of neural networks. Shen et. al delve into a type of attention mechanisms, a recurrent attention and bi-way attention mechansism denoted as a Directional Self-Attention Network (DiSAN). The paper shows that its DiSAN model outperforms complicated RNN models in prediction accuracy and time efficiency on existing benchmarks. The paper is relevant because it presents another, quite new machine learning mechanism that has shown promise for applied machine learning.

The attention mechanism takes advantage of a hidden neural network layer. The hidden layer works on the input sequence and predicts the importance of their weights. This creates a mechanism where neural network inputs are scrutinized by a separate neural network. The separate neural network determines the importance of the weights, and give credence to those weights, so that the model primarily uses the most important inputs. The result is a categorical distribution for the input sequence, and the neural network nodes have memory of which input sequences are important or more relevant. One disadvantage of these networks is that temporal order of input information is lost. The paper's DiSAN model helps fix this by providing sequential memory for the attention networks. The paper demonstrates that their attention mechanism models perform particularly well at alignment scores between two sources, i.e. does well at providing a similarity score between two sources or texts.

There are a few jewels in the Shen et. al paper, like how an additive function for attention often outperforms multiplicative attention, and is also more memory efficient. Their models make us of cross-entropy as an optimization objective and include L2 regularization. The minimization optimizer is Adadelta with mini-batch of size 64. The initial learning rate is quite large, i.e. 0.5, which is decreased over epochs. The weight matrices for networks use GloVe and are pre-trained with out of vocabularty words, which initially were randomely initialized from a uniform distribution. The model uses a dropout of 0.25 and 0.2. The dropout is also varied throughout the learning process. The final model uses fewer parameters than either RNN or CNN networks by margins of 3%. The model is applied to the Standford Sentiment Treebank and performs better than the best existing model by 0.52%. The model is also applied to Sentences Involving Compositional Knowledge (SICK) and with a similar performance. Of important note is the model's bi-directional ability to track different features in forward progressing layers than a backward focused layer, one picking up word families and the latter focusing on word carousel.

Ji Lee performs sequential short-text classification with ANNs. The author's point is that text classifications often occur by only considering a text, not necessarily its preceding or subsequent texts. The paper proposes that using information preceding short texts may improve classification accuracy. Their model initially generates vector representations for short-texts using either the RNN or CNN architectures. The authors utilized early stopping after 10 epochs and performed hyperparameter training. Their model serves as a benchmark for ANN performance to sequential short-text classification.

Wenpeng et al perform a comparative study between CNN and RNN for Natural Langauge Processing (NLP). This is an interseting subject, since RNNs and CNNs differently model sentences. RNNs capture units in sequence and CNNs are good at extracting positional invariant features. Both CNNs and RNNS are also the primary types of DNNs. The paper covers multiple NLP tasks with each type of network, specifically CNNs, Gated Recurrent Units (GRUs), and LSTMs. It is wroth mentioning that the networks in the study did not obtain great performance on existing benchmarks, which may limit the value of the study's insights. The NLP tasks are sentiment/relation classification, textual entailment, answer selection, question-relation matching, and part-of-speech tagging. The authors found that both CNNs and RNNS provide complementary information on text classification tasks. The authors also found that changing hidden layer sizes and batch sizes resulted in large performance fluctuations. A related work in the study found that RNNs compute a weighted sum of n-grams while CNNs extract the most important n-grams and only consider their resulting activation.

I read a paper on sentence pair scoring by Petr et al. The authors argue that many sentence pairing tasks like Answer Set Selection, Semantic Text Scoring, Next Utterance Ranking, and Recognizing Textual Entailment are all very similar. They propose a unified framework that employs task-independent models for sentence pair scoring models. The model can easily compare models against its baselne in an effort to create a better framework for evaluating machine learning models. It could be worthy comparing any models I might create for sentence pair scoring within their model framework.

There were a few papers on deep learning with video data. One interesting paper performed deep learning by using CNNs on mulitple frames. This paper was by Hossein Mobahi. The paper performs large-scale object recognition. Videos are composed of multiple frames, which provides a number of frames which contain the same objects. These similar frames can each be processed by a CNN to provide additional information and possibly better accuracy in object recognition. The paper learns the objects, based on the frame-to-frame video motion by performing classification on each frame. The authors see learning from multiple frames more related to evolution, as humans experience learning through the world, which is constantly moving and changing. The paper made use of 72x72 sized images of 100 images, where each object was shot 100 times at angles that each differed by 5 degrees.

Wojciech Zaremba explores RNN regularization. Dropout is the most successful technique for regularizing neural networks, but they do not work well with RNNS and Long Short-Term Memory units (LSTMs). Dropout works by randomly dropping outputs on a certain percentage of nodes. Dropout is used as a form of regularization to make networks more generic and stable on new inputs. Being able to apply regularization to RNNs or LSTMs could make video deep learning much more performant.

0.3 What area within IT would you like to draw upon to inform a topic that you are interested in?

0.4 Topic

0.4.0.1 What is known about this topic?

0.4.0.2 What needs to be known?

0.4.0.3 What would you like to know?

0.4.0.4 Why is it important?

0.4.0.5 Why we should care?

0.5 Describe your understanding of one or two open problems/questions in IT

0.5.1 What is the gap between what is known and what needs to be known?

0.5.2 What would you like to know?

0.5.3 Formulate a question that specifies b.

0.6 Research Question

0.6.1 Open Research Problems in IT

0.7 Purpose Statement

0.7.1 Signpost that establishes the central intent for the study?

0.7.2 Is there a body of knowledge in IT you would like to contribute to?

0.7.3 Is there an area of practice you would like to improve?

0.8 Theory

0.8.1 Philosophical Assumptions

0.8.1.1 Social Reality

0.8.1.2 What we understand to be true

0.9 How do we understand?

0.10 Means of Investigation

0.10.0.1 Qualitative/Inductive

0.10.0.2 Quantitative/Deductive

0.10.1 Means of Evidence

0.10.1.1 Data Collection

0.10.1.2 Data Analysis

Bibliography

Baudi, P. and Pichl, J. Sentence Pair Scoring : Towards Unified Framework. (C).

Crowe, C. (2018). Initial Related Work. pages 1–4.

Donahue, J., Hendricks, L. A., Guadarrama, S., Rohrbach, M., Venugopalan, S., Saenko, K., Darrell, T., Austin, U. T., Lowell, U., and Berkeley, U. C. Long-term Recurrent Convolutional Networks for Visual Recognition and Description.

Enerative, F. O. R. G., Of, M. O., and Ideos, N. A. V. (2014). V (l) m : a b f g m n v. pages 1–15.

Fan, Y., Lu, X., Li, D., and Liu, Y. (2016). Video-Based Emotion Recognition using CNN-RNN and C3D Hybrid Networks. (November).

Ghosh, A. and Veale, T. (2016). Fracking Sarcasm using Neural Network. pages 161–169.

Graves, A. Generating Sequences With Recurrent Neural Networks. pages 1–43.

Graves, A. (2014). Towards End-to-End Speech Recognition with Recurrent Neural Networks. 32.

Hurri, J. (2003). Simple-Cell-Like Receptive Fields Maximize Temporal. 691(3):663–691.

Ji, S. and Yu, K. (2010). 3D Convolutional Neural Networks for Human Action Recognition.

Karpathy, A., Toderici, G., Shetty, S., Leung, T., Sukthankar, R., and Fei-fei, L. (2015). Large-scale Video Classification with Convolutional Neural Networks Presenter : Esha Uboweja Problem Classification of videos in sports datasets. (June 2014).

Lan, Z., Lin, M., Li, X., Hauptmann, A. G., and Raj, B. Beyond Gaussian Pyramid : Multi-skip Feature Stacking for Action Recognition.

Le, Q. V., Zou, W. Y., Yeung, S. Y., and Ng, A. Y. Learning hierarchical invariant spatio-temporal features for action recognition with independent subspace analysis.

Lee, J. Y. and Dernoncourt, F. (2016). Sequential Short-Text Classification with Recurrent and Convolutional Neural Networks.

M, N. E. N. E. and Yuille, A. (2015). D c m r n n (-rnn). 1090(2014):1–17.

Merri, B. V. and Fellow, C. S. (2014). Learning Phrase Representations using RNN Encoder – Decoder for Statistical Machine Translation. pages 1724–1734.

Michalski, V. and Memisevic, R. Modeling Deep Temporal Dependencies with Recurrent " Grammar Cells ". pages 1–9.

Mobahi, H., Weston, J., America, N. E. C. L., and Way, I. (1996). Deep Learning from Temporal Coherence in Video.

Sainath, T. N., Vinyals, O., Senior, A., and York, N. No Title. pages 1–5.

Simonyan, K. Two-Stream Convolutional Networks for Action Recognition in Videos. pages 1–9.

Simonyan, K. and Zisserman, A. (2015). Very Deep Convolutional Networks for Large-Scale Image Recognition. *International Conference on Learning Representations (ICRL)*, pages 1–14.

Soomro, K., Zamir, A. R., Shah, M., and Recognition, A. (2012). UCF101 : A Dataset of 101 Human Actions Classes From Videos in The Wild. (November).

Srivastava, N. (2014). Unsupervised Learning of Video Representations using LSTMs.

Srivastava, N. (2015). Unsupervised Learning of Video Representations using LSTMs. 37.

Susskind, J., Memisevic, R., Hinton, G., and Pollefeys, M. Modeling the joint density of two images under a variety of transformations.

Sutskever, I. Sequence to Sequence Learning with Neural Networks. pages 1–9.

Understanding, R. N. N. C.-f. L. DiSAN: Directional Self-Attention Network for RNN/CNN-Free Language Understanding.

Vosoughi, S. and Roy, D. (2016). Tweet2Vec : Learning Tweet Embeddings Using. pages 16–19.

Wang, J., Yang, Y., Mao, J., Huang, Z., Huang, C., and Xu, W. CNN-RNN : A Unified Framework for Multi-label Image Classification. pages 2285–2294.

Wang, J., Yu, L.-c., Lai, K. R., and Zhang, X. (2016). Dimensional Sentiment Analysis Using a Regional CNN-LSTM Model. pages 225–230.

Yin, W., Kann, K., and Yu, M. (2016). Comparative Study of CNN and RNN for Natural Language Processing.

Yin, W. and Sch, H. ABCNN: Attention-Based Convolutional Neural Network for Modeling Sentence Pairs.

Zaremba, W. and Com, V. G. (2013). arXiv : 1409 . 2329v3 [cs . NE] 3 Nov 2014.