Reading your paper:  
second pass

*Submitted by Abdulrahman Alabrash*

# OVERVIEW & PURPOSE

Reading a technical paper is unlike reading fiction: prepare to spend a lot more time, do multiple passes, take notes, survey literature, figure out the math, understand the algorithms. You are just starting machine learning, so many concepts will be unfamiliar. However you will have a whole semester with your paper and we will follow a process in which your understanding will improve gradually.

# HOW TO READ

To help you with this task please first read a couple of guides on how to read:

* [How to read a paper](http://www.sigcomm.org/sites/default/files/ccr/papers/2007/July/1273445-1273458.pdf) by S. Keshav.
* [How to read a technical paper](https://www.cs.jhu.edu/~jason/advice/how-to-read-a-paper.html) by J. Eisner.

# SECOND PASS

Go over your paper more carefully using Keshav’s second pass. This should take you about an hour. Please give the reference to your paper and answer the questions below. The paper reference should have the format “Author, Year, Title, Howpublished” (you can optionally include a link).

**Paper reference:**

1. Data:Summarize the datasets used in the experimental section of your paper. Indicate details like number of instances, number of features, number of sentences/words (if text), dimensions (if image), etc.

Three benchmark datasets in two categories was used ,

**Sentiment analysis,**

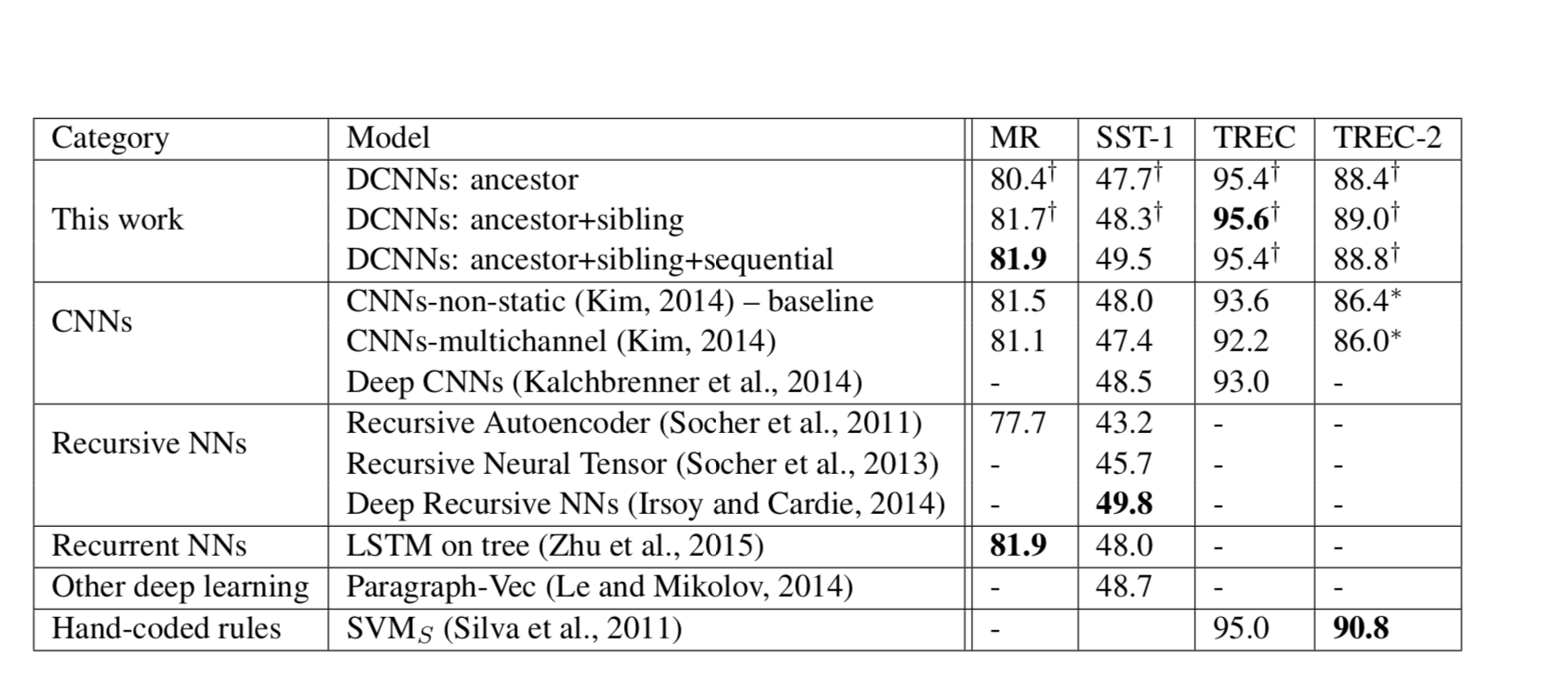
10,662 sentences in the Movie Review (MR ) (Pang and Lee 2205) each instance is labeled positive or negative and in most cases contains one sentence. Since no standard data split is given 10 fold cross validation is used to include every sentence in the training and testing at least once. The final model outperforms the baseline sequential CNNs by 0.4

SST (Stanford Sentiment Treebank) (Schoer et al 2013) dataset is different from MR, SST annotates finer-grained labels, very positive, positive, neutral, negative and very negative. There are 11,855 sentences with standard split. The model achieves an accuracy of 49.5

**Question classification** on TREC, the entire dataset of 5,952 sentences are classified into following 6 categories: abbreviation, entity, description, location, and numeric, in this experiment.

1. Baselines: Does your paper give any baselines? Baselines are the simplest models you can imagine that give you a lower bound on the performance. In a classification algorithm a baseline can be picking your class randomly, or always picking the most common class. In a regression problem a baseline can be to always choose the average of all the outputs in the training set.

a table of result is shown to compare the model of DCNN with different version with other models.



1. Upper bounds: Does your paper give any upper bounds on the possible performance? 100% accuracy may not always be possible if outputs are noisy, or if the inputs are incomplete, (i.e. the same input may lead to different outputs in different instances) etc. For example, in word sense disambiguation or machine translation two people do not always agree on what the best answer is (*inter-annotator agreement* is a commonly used metric to measure human agreement level on a task). It would be unreasonable to expect a computer model to achieve 90% accuracy on a task where humans agree only 70% of the time.

There is no upper bound given in the paper.

1. State-of-the-art: What is the state-of-the-art on the task(s) studied in your paper? These are the best published results on the task. How much did your paper improve the state-of-the-art? Did you run into any papers in your future/related work research that improved it any further?

* Achieve the highest published accuracy on TREC, In this experiment, DCNNs easily outperform any other methods even with ancestor convolution only. DCNNs with sibling achieve the best performance in the published literature. DCNNs combined with sibling and sequential information might suffer from overfitting on the training data based on our observation.
* DCNNs with sibling, Ancestor paths alone is not enough to capture many linguistic phenomena such as conjunction. many linguistic phenomena such as conjunction. Inspired by higher-order dependency parsing, in various ways.
* No.

1. Ablation study: What are the different ideas that combined to give the performance reported in the paper? How important is each of them? Did the authors describe what happens when you remove these ideas one at a time (this is known as an ablation study)? Did the authors start with a baseline system and add these ideas one at a time to show you what the impact is? If the paper gives you a combination of N ideas and a single result, it is a bad paper: you have no idea which of these N ideas is more important.

There is no ablation study is mentioned on the paper.

# UNKNOWN TERMS

Take the list of unknown terms you have prepared in the last assignment and look for definitions on the web. You can use the [deep learning glossary](http://www.wildml.com/deep-learning-glossary/) ([here are some more](https://www.google.com.tr/search?q=deep+learning+glossary)), [deeplearning.net](http://deeplearning.net/), a [textbook](http://www.deeplearningbook.org/), [wikipedia](https://en.wikipedia.org/wiki/Main_Page), [scholarpedia](http://www.scholarpedia.org/article/Main_Page), etc. (Please let me know if you find any other resources so I can add them to the list).

* Sequence labeling, is a type of pattern recognition task that involves the algorithmic assignment of a categorical label to each member of a sequence of observed values. A common example of a sequence labeling task is part of speech tagging, which seeks to assign a part of speech to each word in an input sentence or document. Sequence labeling can be treated as a set of independent classification tasks, one per member of the sequence. However, accuracy is generally improved by making the optimal label for a given element dependent on the choices of nearby elements, using special algorithms to choose the globally best set of labels for the entire sequence at once.
* Semantic parsing, is the process of mapping a natural-language sentence into a formal representation of its meaning. A shallow form of semantic representation is a case-role analysis (a.k.a. a semantic role labeling), which identifies roles such as agent, patient, source, and destination. A deeper semantic analysis provides a representation of the sentence in predicate logic or other formal language which supports automated reasoning.
* Search query retrieval, is a new Convolutional Deep Structured Semantic Models (C-DSSM). Compared with DSSM, C-DSSM has a convolutional layer that projects each word within a context window to a local contextual feature vector. The model has been developed with Microsoft.
* N-grams tree, is a contiguous sequence of n items from a given sample of text or speech. The items can be phonemes, syllables, letters, words or base pairs according to the application. The n-grams typically are collected from a text or speech corpus. When the items are words, n-grams may also be called shingles.
* CNN (Convolution, filter, and pooling operations ), Conventual neural network has been explained by Deniz professor last week, CNNs use a variation of multilayer perceptrons designed to require minimal preprocessing. They are also known as shift invariant or space invariant artificial neural networks (SIANN), based on their shared-weights architecture and translation invariance characteristics.

# RELATED WORK

Go over the paper’s related work, both past and future, you found in the last assignment. Apply [Keshav’s first pass](http://www.sigcomm.org/sites/default/files/ccr/papers/2007/July/1273445-1273458.pdf) to each, this should take 5-10 minutes per paper. Please give the citations and your understanding of these papers (and whether/why they matter to your paper) below:

* Citation1: Goldberg, Yoav. "A primer on neural network models for natural language processing." Journal of Artificial Intelligence Research 57 (2016): 345-420
* Category: Tutorial surveys, prime to Natural language processing using Deep learning tools
* Context: This tutorial surveys neural network models from the perspective of natural language processing research, in an attempt to bring natural-language researchers up to speed with the neural techniques.
* Correctness: it is citied over 300 times which indicates its validity.
* Contributions: it is aimed at those readers who are interested in taking the existing, useful technology and applying it in useful and creative ways to their favourite NLP problems. For more in-depth, general discussion of neural networks,
* Clarity: the paper written was very clear
* Citation2: Lin, Zhouhan, et al. "A structured self-attentive sentence embedding." arXiv preprint arXiv:1703.03130 (2017).
  + Category: description of a new approach for sentence embedding.
  + Context: introducing a new model of sentence embedding using 2-D matrix instead of using vector. Experimental results over 3 different tasks show that the model outperforms other sentence embedding models by a significant margin.
  + Correctness: despite the paper is published in 2017 but the its cited number over 200. Which indicate its trusted by a lot of researcher, and its valid.
  + Contributions: the paper introduced a fixed size, matrix sentence embedding with a self-attention mechanism. Because of this attention mechanism, there is a way to interpret the sentence embedding in depth in the model.
  + Clarity: the paper is well written.
* Citation3: Goldberg, Yoav. "Neural network methods for natural language processing." Synthesis Lectures on Human Language Technologies 10.1 (2017): 1-309.
  + Category: it’s a review of Neural Network Methods for Natural Language Processing book.
  + Context: a brief summary for the 4 parts of the book.
  + Correctness: the assumptions appear to be valid.
  + Contributions: the paper explain how the book provides valuable materials for newcomers into this exciting arena of cross-disciplinary research, by preparing relevant information of both neural networks and natural language processing
  + Clarity: it is well written.