DEEP LEARNING FOR DOCUMENT CLASSIFICATION

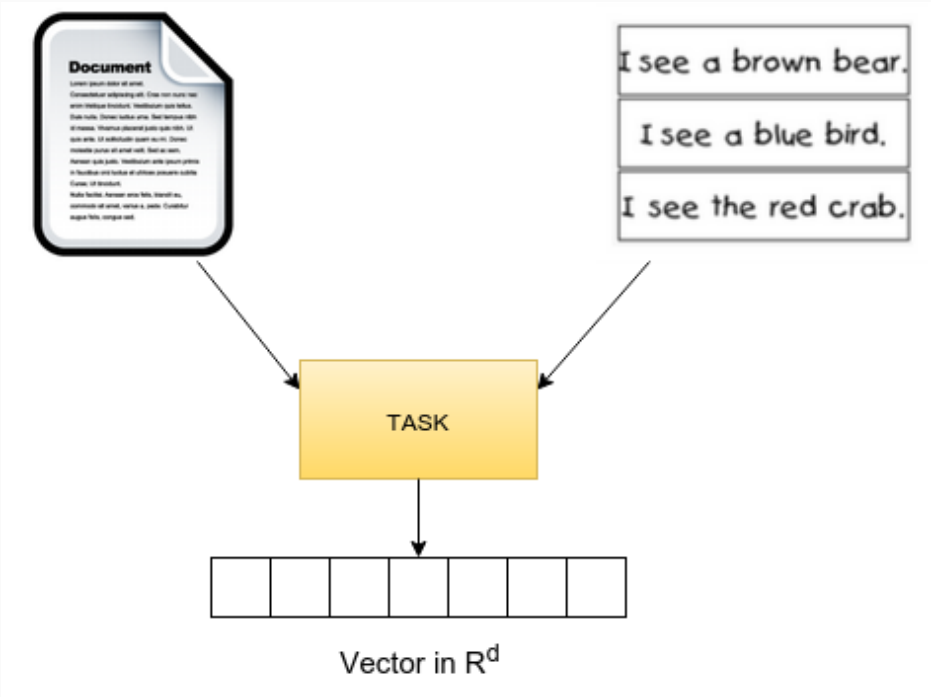
CS671 - Course Project

Amlan Kar
Sanket Jantre
Mentored by: Prof. Amitabha Mukerjee

Indian Institute of Technology, Kanpur

Motivation

- **Creation** and usage of new *task-specific* Sentence and Word level vectors for efficient semantic representation for application in Document and Sentence Classification tasks.
- Results in (Y.Kim, EMNLP 2014)[1] show promise and scope.



Why Deep Learning ?

- Breaking State of the Art barriers in computer vision (Krizhevsky et al., 2012) and speech recognition (Graves et al., 2013),
- Recent advances in standard NLP tasks have all come through the application of Deep Learning in tandem with Statistical Methods in ensemble learners.

Why Convolutional Neural Networks ?

- Possibility of parse-tree like feature graphs (by looking at the firing neurons) that show induced non-linear composition used for classification in NLP tasks.

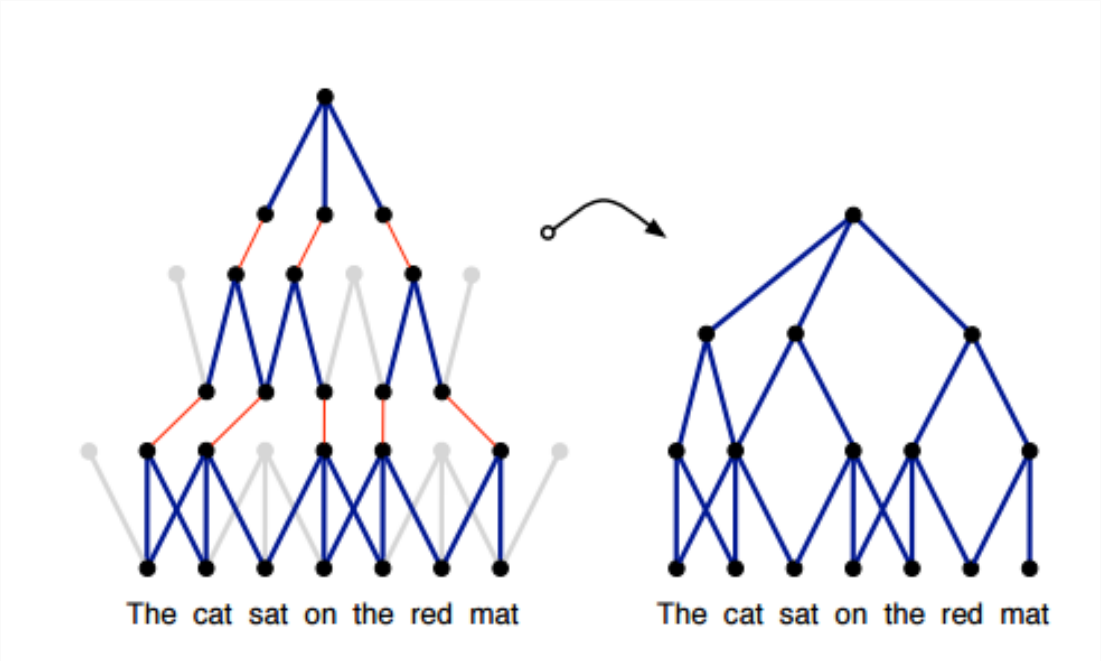


Figure: Image from (Kalchbrenner et al., 2014) [2]

Approach

We plan to model our sentence or document as a 2D matrix using word2vec embeddings[3] of words for sentences and Skip-Thought embeddings[4] of sentences for documents.

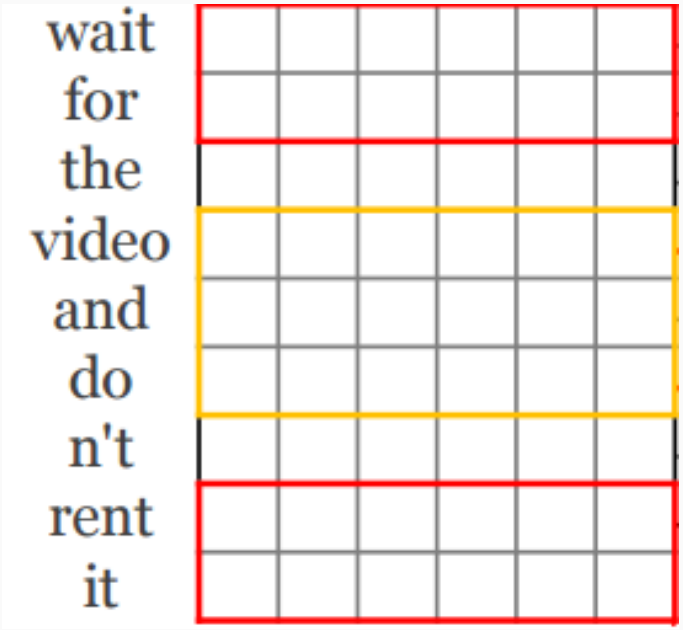


Figure: Image from (Y.Kim, 2014) [1]

Approach

Static Channel: The case where we treat the word vectors as static input.
Non-Static Channel: The case where we fine-tune the word vectors during training.

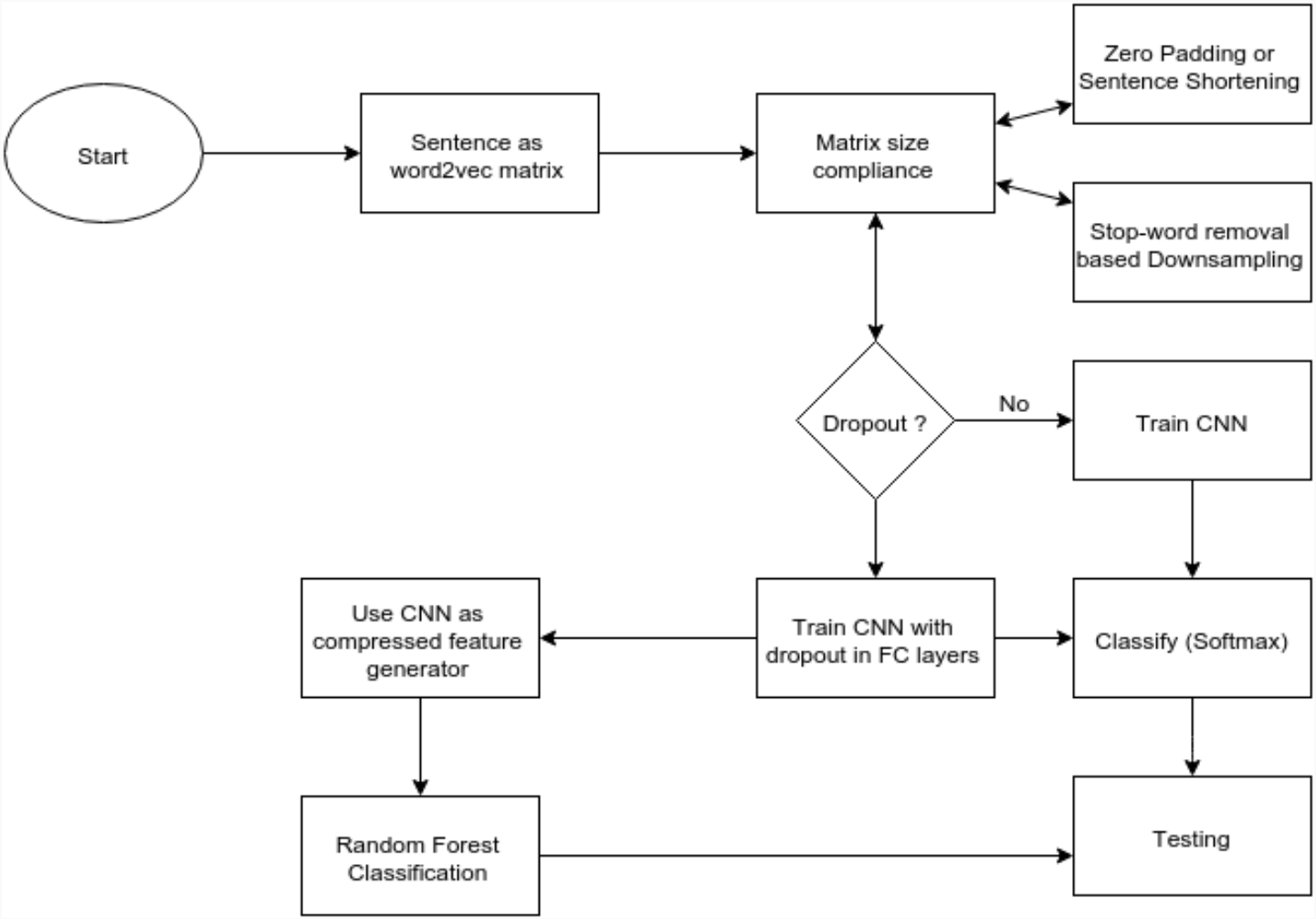
Rationale: The Non-Static channel method has been shown to generate much better semantic embeddings[1]. It also seems natural, as we humans seem to apply domain specific knowledge to a general model while solving a specific problem. Why not have domain specific fine-tuned vectors?

Approach

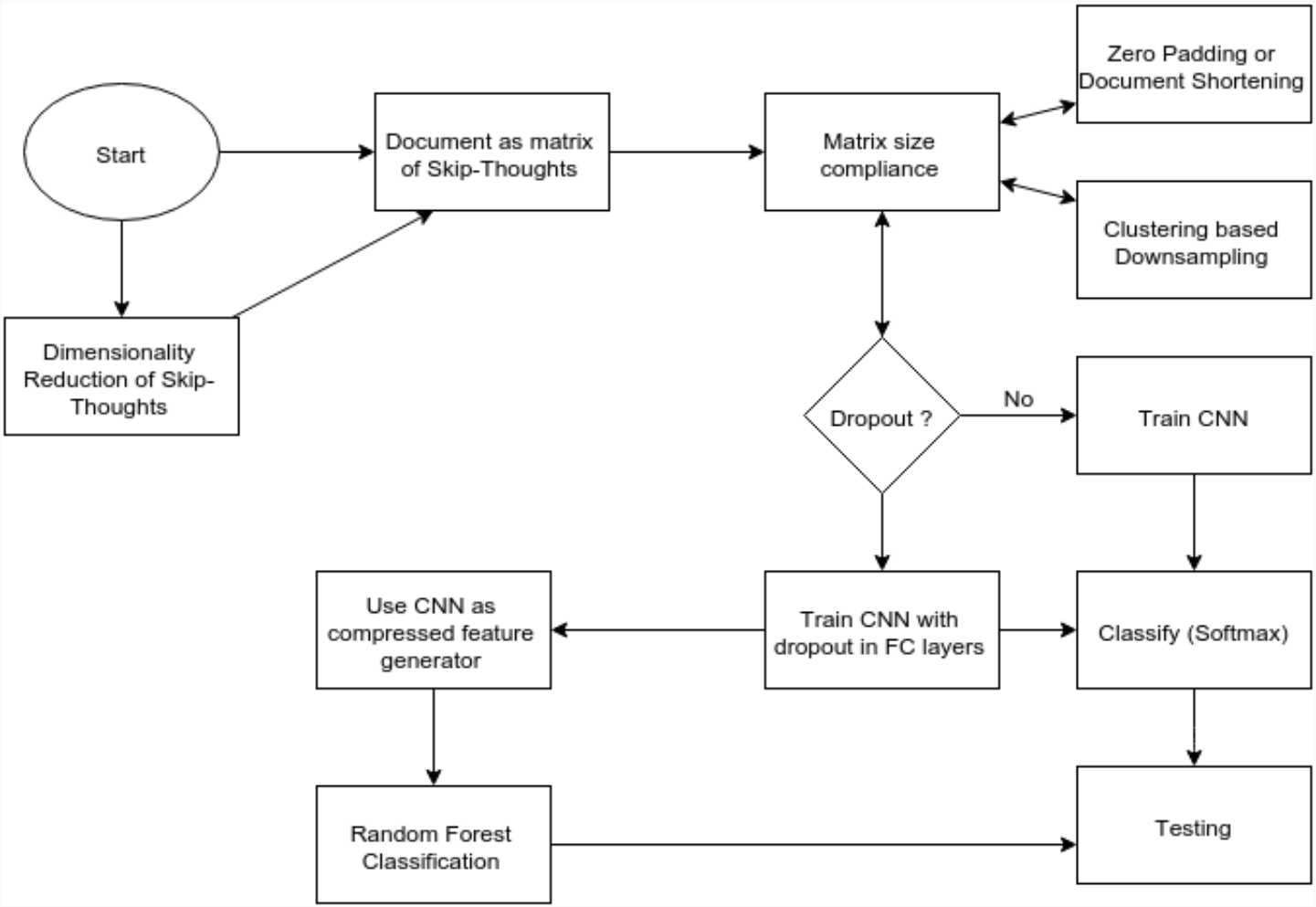
	Most Similar Words for	
	Static Channel	Non-static Channel
bad	*good* *terrible* *horrible* *lousy*	*terrible* *horrible* *lousy* *stupid*
good	*great* *bad* *terrific* *decent*	*nice* *decent* *solid* *terrific*
n't	*os* *ca* *ireland* *wo*	*not* *never* *nothing* *neither*

Figure: Image from (Y.Kim, 2014) [1]

Approach - Sentence



Approach - Document



ConvNet Structure

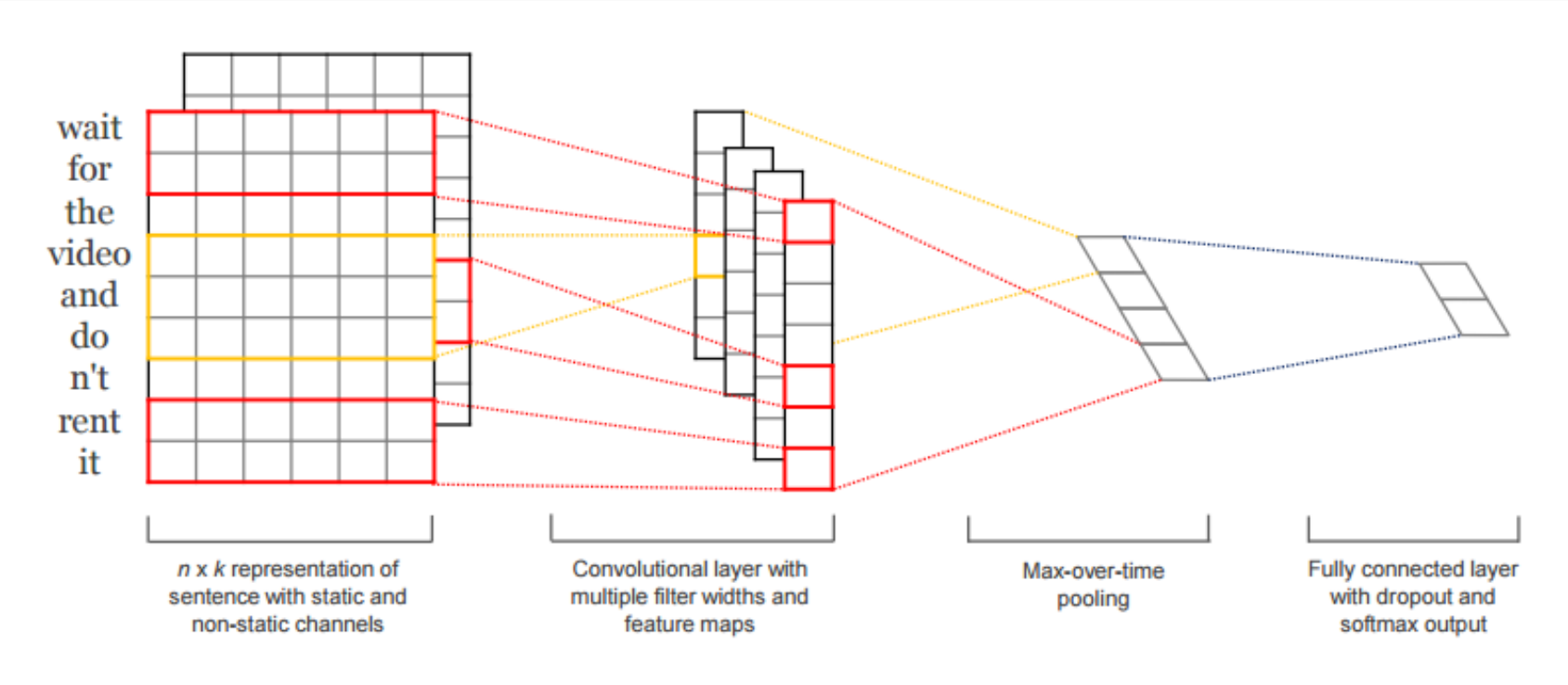


Figure: Multi-channel ConvNet[1]

ConvNet Structure

Our ConvNet structure is slight variant of the one proposed by Collobert et al. (2011)[5] and similar to the one used by Kim. (2014)[1].

- We propose to employ wide-convolution instead of simple convolution that was used by Y.Kim.
- We will do a k-max-over-time pooling instead of normal max-over time pooling and concatenate to get the FC-1 layer input.

Work Done

- Datasets collected for various core NLP tasks.
- ConvNet code almost complete.
- Implementation Details
  - Code has been written in Python using the Theano deep learning library and the Keras library.
  - Mini-batch SGD is used for backpropagation.
  - We will use both a ReLU and a tanh non linearity and compare.
  - Dropout is being used in the Fully connected layer to prevent co-adaptation of features.
  - Word vectors are obtained from Google's trained model on the Google News dataset.
  - Skip-thought vectors are obtained from the RNN encoder-decoder model released by Ryan Kiros.

Future Work

- We intend to try and fine-tune phrase vectors if this work gets done in time. For this, we intend to use Collobert's Senna software for phrase chunking before vector production by composition on word-vectors as suggested by Mikolov et al.[3].

- Train word2vec on a Hindi corpus before employing this method on the Hindi Movie Review sentiment classification task.

- We also wish to try out this method on Multi-class document classification which is a field that has not been touched significantly by the deep learning revolution yet.

Done!

References I


Yoon Kim.
Convolutional neural networks for sentence classification.
EMNLP 2014, 2014.


Nal Kalchbrenner, Edward Grefenstette, and Phil Blunsom.
A convolutional neural network for modelling sentences.
arXiv preprint arXiv:1404.2188, 2014.


Tomas Mikolov, Ilya Sutskever, Kai Chen, Greg S Corrado, and Jeff Dean.
Distributed representations of words and phrases and their compositionality.
In *Advances in neural information processing systems*, pages 3111–3119, 2013.

References II

Ryan Kiros, Yukun Zhu, Ruslan Salakhutdinov, Richard S Zemel, Antonio Torralba, Raquel Urtasun, and Sanja Fidler.
Skip-thought vectors.
arXiv preprint arXiv:1506.06726, 2015.

Ronan Collobert, Jason Weston, Léon Bottou, Michael Karlen, Koray Kavukcuoglu, and Pavel P. Kuksa.
Natural language processing (almost) from scratch.
CoRR, abs/1103.0398, 2011.