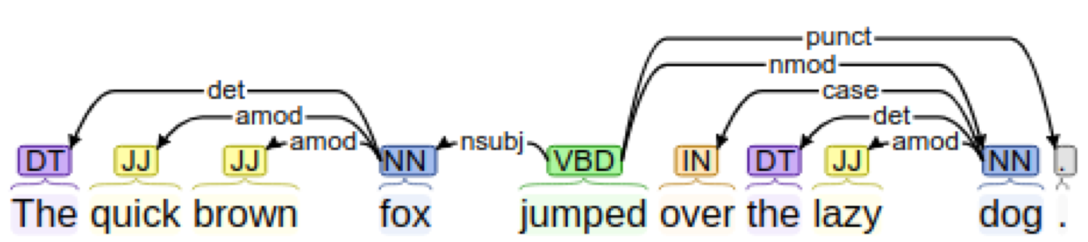
Incorporating Syntactic and Semantic Information in Word Embeddings using Graph Convolutional Network

Shikhar Vashishth[1], Manik Bhandari[2*], Prateek Yadav[3*], Piyush Rai[4], Chiranjib Bhattacharyya[1], Partha Talukdar[1]
Indian Institute of Science, Carnegie Mellon University, Microsoft Research, and IIT Kanpur

Motivation

Word Embeddings have been widely adopted across several NLP applications. However, most of the existing methods utilize sequential context of a word to learn its representation.

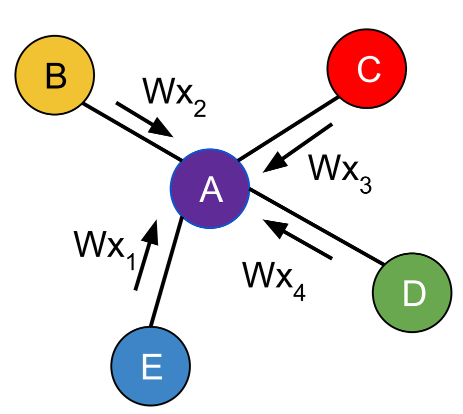
In this work, we explore to utilize syntactic context of words for learning word embeddings using recently proposed Graph Convolutional Networks. Also, we propose a more effective way for incorporating semantic knowledge like *synonyms, hyponyms in learned embeddings.*



Graph Convolutional Networks

GCNs are a generalization of Convolutional Neural Networks for non-Euclidean data. In this work, we utilize first-order approximation of GCNs (Kipf et al. 2016). The update equation for node A in the graph is given as:

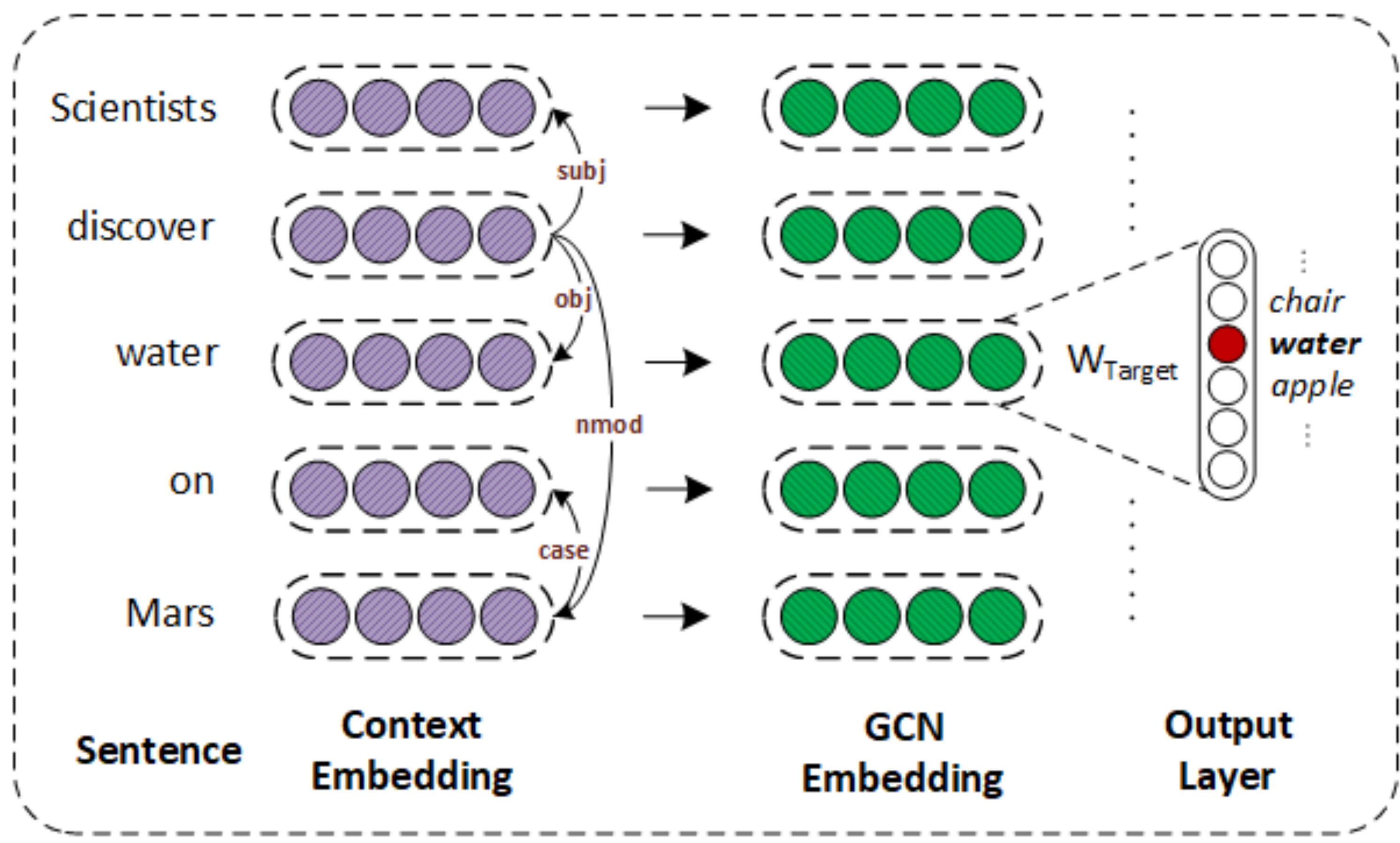
$$h_v = f\left(\sum_{u \in \mathcal{N}(u)} W x_u + b\right).$$



Contributions

1. We propose SynGCN, a Graph Convolution based method for learning word embeddings. Unlike the previous methods, SynGCN utilizes syntactic context for learning word representations without increasing vocabulary size.
2. We also present SemGCN, a framework for incorporating diverse semantic knowledge in learned word embeddings, without requiring relation-specific special handling as in previous methods.
3. Through experiments on multiple intrinsic and extrinsic tasks, we demonstrate substantial improvement over state-of-the-art approaches, and also yield an advantage when used in conjunction with methods such as ELMo.

SynGCN Overview



SynGCN (Sentence-level)

For each word in the sentence, SynGCN learns its representation by aiming to predict the word based on its dependency context encoded using GCNs defined as:

$$h_i^{k+1} = f\left(\sum_{j \in \mathcal{N}(i)} g_{l_{ij}}^k \times \left(W_{l_{ij}}^k h_j^k + b_{l_{ij}}^k\right)\right) \qquad h_i = \sum_{-c \le j \le c, j \ne 0} h_j,$$

Proposed method (SynGCN) **CBOW**

Results

SynGCN Evaluation: Performance comparison on multiple intrinsic and extrinsic tasks. Overall, we observe that SynGCN outperforms all the existing word embeddings methods.

	Word Sim	Concept Cat	Analogy	POS	SQuAD	NER	Coref
Word2vec	71.4	63.2	44.0	95.0 ± 0.1	78.5 ± 0.3	89.0 ± 0.2	65.1 ± 0.3
GloVe	69.2	58.0	45.8	94.6 ± 0.1	78.2 ± 0.2	89.1 ± 0.1	64.9 ± 0.2
Deps	65.7	61.8	40.3	95.0 ± 0.1	77.8 ± 0.3	88.6 ± 0.3	64.8 ± 0.1
EXT	69.6	52.6	18.8	94.9 ± 0.2	$\mathbf{79.6 \pm 0.1}$	88.0 ± 0.1	64.8 ± 0.1
SynGCN	**73.2**	**69.3**	**52.8**	$\mathbf{95.4 \pm 0.1}$	$\mathbf{79.6 \pm 0.2}$	$\mathbf{89.5 \pm 0.1}$	$\mathbf{65.8 \pm 0.1}$

SemGCN Evaluation: Evaluating incorporation of sematic information. M(X, R) denotes the fine-tuned embeddings using method M taking X as initialization embeddings and R type of semantic relations. SemGCN with SynGCN gives the best performance overall.

	Word Sim	Concept Cat	Analogy	POS	SQuAD	NER	Coref
X = SynGCN	61.7	69.3	52.8	95.4 ± 0.1	79.6 ± 0.2	89.5 ± 0.1	65.8 ± 0.1
Retro-fit (X,1)	61.2	67.1	51.4	94.8 ± 0.1	79.6 ± 0.1	88.8 ± 0.1	66.0 ± 0.2
Counter-fit (X,2)	55.2	66.4	31.7	94.7 ± 0.1	79.8 ± 0.1	88.3 ± 0.3	65.7 ± 0.3
JointReps (X,4)	60.9	68.2	24.9	95.4 ± 0.1	79.4 ± 0.3	89.1 ± 0.3	65.6 ± 0.1
SemGCN (X,4)	**65.3**	**69.3**	**54.4**	$\mathbf{95.5 \pm 0.1}$	$\mathbf{80.4 \pm 0.1}$	$\mathbf{89.5 \pm 0.1}$	$\mathbf{66.1 \pm 0.1}$

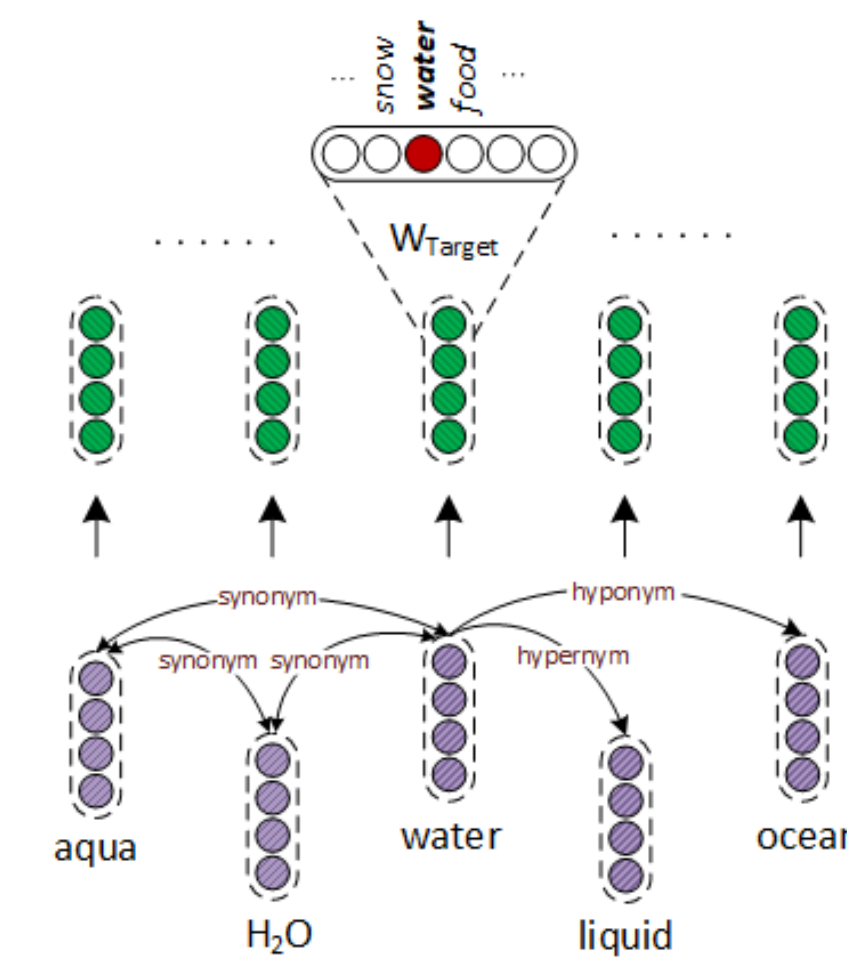
SemGCN Overview

SemGCN allows to incorporate both symmetric (e.g. synonyms) and asymmetric (e.g. hypernyms) semantic knowledge into learned word embeddings. Unlike SynGCN, SemGCN operates on a corpus-level directed labeled graph.

Formally, we aim to maximize the following in both the models:

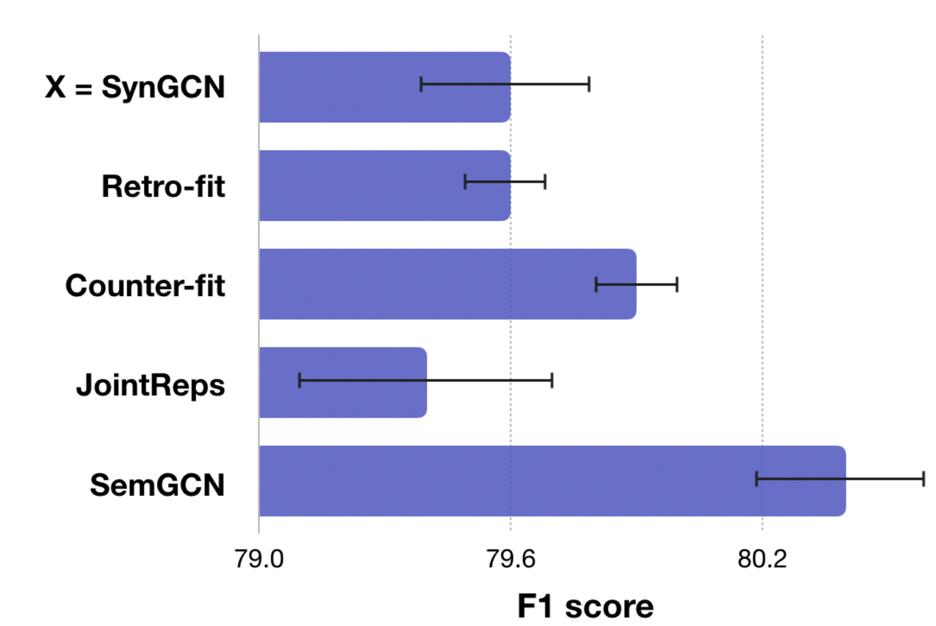
$$E = \sum_{t=1}^{|V|} \left(v_{w_t}^T h_t - \log \sum_{i=1}^{|V|} \exp(v_{w_i}^T h_t) \right),$$

where h_t is the GCN representation of the target word w_t and v_{w_t} is its target embedding.

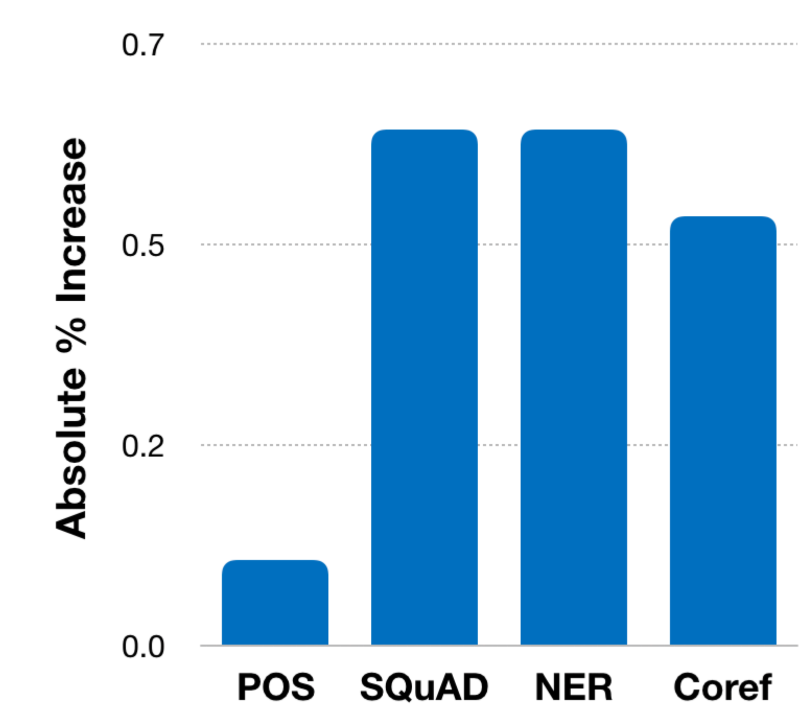


SemGCN (Corpus-level)

Ablation Results and Performance with ELMo



SemGCN gives considerable improvement on SQuAD dataset compared to other methods when provided with the same semantic information (*synonyms*) for fine tuning SynGCN embeddings.



Comparison of ELMo with SynGCN and SemGCN embeddings on multiple extrinsic tasks. Our syntactic embeddings capture complementary information which is not captured by ELMo.

Acknowledgement

This work is supported in part by the Ministry of Human Resource Development (Government of India) and Google PhD Fellowship.

Source Code

Source code is available at: ***github/malllabiisc/WordGCN***

Contact: *shikhar@iisc.ac.in*