

# Recognizing Textual Entailment based on Deep Learning Approach

Mohamed H. Haggag  
Computer Science Department  
Faculty of Computers and  
Information  
Helwan University,  
Cairo, Egypt

Marwa M. A. ELFattah  
Computer Science Department  
Faculty of Computers and  
Information  
Helwan University,  
Cairo, Egypt

Ahmed Mohammed Ahmed  
Ministry of Communication and  
Information Technology  
Telecom Egypt Company, Egypt  
Giza, Egypt

## ABSTRACT

Textual entailment (TE) is a relation that holds between two pieces of text where one reading the first piece can conclude that the second is most likely true. This paper proposes new model based on deep learning approach to recognize textual entailment. The deep learning approach is based on syntactic structure [Holder- Relation - Target] [1] which contains all lexical, syntactic and semantic information about the input text. The proposed model constructs deep leaning neural networks, which aims at building deep and complex encoder to transform a sentence into encoded vectors. The experimental results demonstrate that proposed technique is effective to solve the problem of textual entailment recognition

## General Terms

Textual entailment, Deep learning

## Keywords

Keywords are your own designated keywords which can be used for easy location of the manuscript using any search engines.

## 1. INTRODUCTION

Textual entailment is defined as a relationship between a coherent text T and a language expression, which is considered as a hypothesis, H. it was said that T entails H (H is a consequent of T), if the meaning of H, as interpreted in the context of T, can be inferred by a human from the meaning of T.

A spectrum of approaches has been proposed for Recognizing Textual Entailment (RTE). Most of RTE systems are based on Machine Learning, lexical or semantic approaches [2]. However, the entailment decision problem can be considered as a classification problem. Such systems use features such as lexical, syntactic and semantic features.

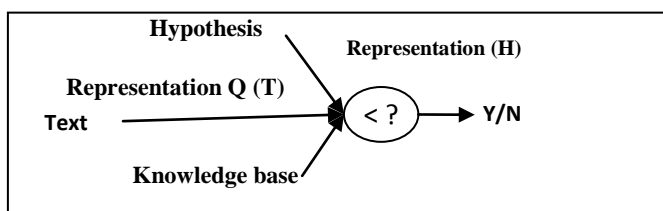


Fig 1: Textual entailment recognition

The typical approach for recognizing textual entailment is usually made up of three main components: a representation component, a comparison component, and a decision component. Some models add pre-processing component

before representation component.

- The representation component involves the representation of the text and the hypothesis in a way to facilitate the comparison between the two. A text can be represented as a tree, a set of predicate argument structures, a logical form or with other ways of representations.
- The comparison component compares the representation of the text and hypothesis which depends on text representation such as using semantic similarity for comparing between the text and hypothesis.
- The decision component is used to decide if the text is entailed the hypothesis or not depending on the comparison component output

Deep Neural Networks (DNNs) are extremely powerful machine learning models that achieve excellent performance on difficult problems such as speech recognition and visual object recognition. DNNs are powerful because they can perform arbitrary parallel computation for text recognition. DNN aims at building deep and complex encoder to transform a sentence into encoded Vectors. Those kinds of neural networks that convert the input data into different representation vectors is called an encoder. The encoder is trained to preserve as much information as possible when an input is run through it, and also to make the new representation have various nice properties.

The paper proposes a new model that builds new syntactic structure for text and hypothesis. Then, this structure is used with Deep neural network to determine if the text is entailed the hypothesis or not. Deep learning neural network constructs from one hundred hidden layers. Experiments results show the contribution of the idea on textual entailment performance.

## 2. RELATED WORK

Earlier work in text entailment was based on lexical analysis only, as Adams approach [3], which is example of lexical approach. The comparison component of this method operates on a combination of word similarity measures, a web based word similarity method, and the lexical edit distance for comparing T and H. The word similarity method used is the method that uses lexical database relations as a similarity measurement. Oren model [4] is a web method that is based on web frequencies to count similarities which are used in. The lexical edit distance simply counts the number of words that were not identical in H and T relative to the length of H, which is seen as insertion from an editing perspective. Then the computed measurements are used as an input to decision

tree classifier. The classifier decides whether T is entailed on H or not. The approach achieved a relatively high accuracy of 0.63 on the RTE2 challenge.

Set of hybrid approaches used a combination of methods to recognize textual Entailment. Hybrid approaches are usually based on only two methods with one acting as primary strategy and the other as a backup. However, some are based on multiple methods with voting mechanism to select the best result. Bos's approach [5] combines two modules, one based on a bag of words and the other based on logical representation and inference. For the first method word overlap and word weight that are calculated as the inverse document frequency from the web are used to compute relatedness.

The new trends on text entailments approaches are built on Knowledge Base Completion, as Masashi Yoshikawa[6] model shows that the processing time of a state-of-the-art logic-based RTE system can be significantly reduced by replacing its search-based axiom injection (abduction) mechanism by that based on Knowledge Base Completion (KBC). It integrates this mechanism in a Coq plug-in that provides an automatic proof tactic for natural language inference. Additionally, it shows that adding new knowledge data contributes to better RTE performance while not harming the processing speed in this framework. It presents an approach to axiom injection, which, by not holding databases explicitly, allows handling of massive amount of knowledge without losing efficiency. It is built on Knowledge Base Completion (KBC), which recently has seen a remarkable advancement in the machine learning community.

Qian Chen's model [7] enriches neural-network-based NLI (Natural language inference) models with external knowledge in co-attention, local inference collection, and inference composition components. Qian Chen's model achieved better performances on the SNLI and Multi-NLI datasets. Using external knowledge is more significant when the size of training data is restricted, as more knowledge can be obtained.

Chaitanya's model [8] creates (for experimental purposes) an entailment dataset for the clinical domain, and a highly competitive supervised entailment system, called ENT. The model explored self-training and active learning strategies to address the lack of labeled data. Their ENT is used as the basis of self-training and active learning experiments.

Deep learning neural networks which convert the input data into different representation vectors is called an encoder[9]. The encoder is trained to preserve as much information as possible when an input is run through it, and also to make the new representation have various nice properties. One from these encoders was Chen Lyu [10] model, which is built on Restricted Boltzmann Machines to learn the joint representation. It aims to capture the shared meaning underlying the T and H. Recognition approaches based on vector space models of semantics.

### 3. THE PROPOSED TEXT ENTAILMENT MODEL

The proposed model is based on hybrid approach which is based on lexical, syntactic, semantic analysis and Deep Learning classifier entailment module. Stanford CoreNLP is used for tokenization, stem, part of speech tagging, and identify semantic roles of sentences for texts and hypothesis. Depending on the linguistic structure levels for which syntactic information is provided, syntactic parsing might be divided into shallow parsing and fully syntactic parsing. The Stanford parser [11] provides dependency tree as well as phrase structure trees. The Stanford Dependency Parser has been run for extracting dependency relations obtained for a text and hypothesis pair.

Semantic analysis on proposed model based on building new Syntactic structure [Holder- Relation - Target] for text and hypothesis. This syntactic structure is used to represent text to train deep learning classifier. The structure of deep learning model is shown as the following.

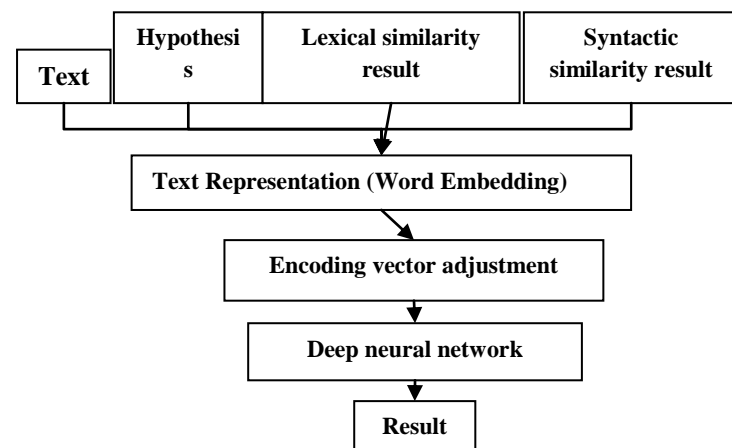


Fig 2: Textual entailment recognition proposed model

The first step in the proposed model is text representation for input data. This step aims to numerically represent the structure data for the input text and hypothesis to make them mathematically computable. There are multiple ways to encode a text such as Co-occurrence count or using neural networks.

#### 3.1 Text Representation

A fundamental problem of text mining is how to represent the text data to make them mathematically computable. Various text representation strategies have been proposed in the past decades for different application purposes such as text categorization, novelty detection and Information Retrieval (IR). Distributed representations of words (or word embeddings) are used to provide useful features for various tasks in natural language processing.

Representing text data has attracted much attention of researchers in different areas due to its great industrial and commercial application potentials. The proposed representation method is based Syntactic structure [Holder-Relation - Target] for text and hypothesis and lexical and syntactic similarity results which came from lexical analysis module and syntactic similarity module.

<p><u>Complements</u></p> <p>Adverb of place                  Adverb of time                  Adverb                  Adjective</p>	<p><u>Target</u></p> <p>Object                  Sentence                  Paragraph</p>	<p><u>Relation</u></p> <p>Sentence                  Event                  Connector</p>
<p><u>Holder</u></p> <p>Object                  Sentence                  Paragraph</p>		

<p><u>Complements</u></p> <p>1-Type                  2- POS                  3-Target Syntactic Type                  4- Parent Syntactic Type                  5-<u>Sense</u></p>	<p><u>Target</u></p> <p>1-Type                  2- POS                  3-Target Syntactic Type                  4- Parent Syntactic Type                  5-<u>Sense</u></p>	<p><u>Relation</u></p> <p>1-Type                  2- POS                  3-Relation Syntactic Type                  4- Parent Syntactic Type                  5-<u>Sense</u></p>
<p><u>Holder</u></p> <p>1-Type                  2- POS                  3-Holder Syntactic Type                  4- Parent Syntactic Type                  5-<u>Sense</u></p>		

**Fig 3. Holder- Relation-Target structure**

The input vector length to text and hypothesis are based on Holder-Relation-Target structure .the length of each part in input vector is shows in next table.

**Table 1. Holder-Relation-Target structure length**

Object Name	Object Size
Holder object	58 unit
Relation object	58 unit
Target object	58 unit
Complements	290 unit
Total Size	464 unit

### 3.2 Encoding vector adjustment

Sometimes the semantic information for the text exceeds the specified vector length for the text so the proposed model adjusts the semantic information length to match the specified vector length depending on semantic analysis for the input text.

### 3.3 Train deep learning neural network

After encode the syntactic information for input text and Hypothesis, these vectors is used to train the deep learning neural network. Here the structure for the sentence “**Fiber improves blood sugar control.**” The Holder-Target-Relation structure is



**Fig 4. Holder- Relation-Target structure for sentence.**

## 4. EXPERIMENTS

### 4.1 Dataset

In conducted experiments, the labeled texts and hypothesis pairs provided by the Text Analysis Conference (TAC) are used. The data sets RTE1\_DEV, RTE2\_DEV and RTE3\_DEV are used as the training data and RTE1\_TEST, RTE2\_TEST and RTE3\_TEST are used as the testing data. Table 2 shows the size of each data set.

**Table 2: Data set Size**

Data Set Name	Data Set Size
RTE1_DEV	567
RTE2_DEV	800
RTE3_DEV	800
RTE1_TEST	800
RTE2_TEST	800
RTE3_TEST	800
Total	4567

### 4.2 Model accuracy

The proposed model uses 1000 samples to test the proposed model. The next table shows the result of proposed model

**Table 3: Test Set Details**

Group	True	False	Total
Positive	533	67	600
Negative	365	35	400

The proposed model uses these Data Sets for generating specific Data Sets. It divides this data into groups each group has Train Data and test data. Then it tried to make these groups varying in data size. It uses these groups for measuring the learning accuracy for proposed model and evaluating proposed model. The following table shows the data size of each group.

**Table 4: Model evaluation**

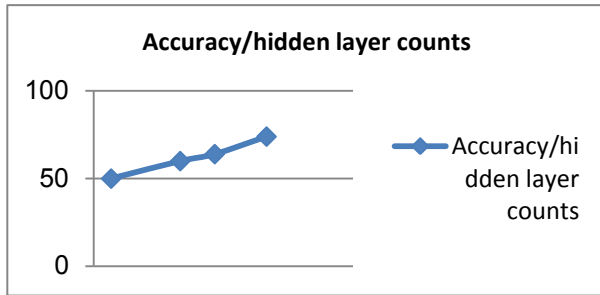
Group Name	Training Data Size	Testing Data Size	Total group size
Group A	50	10	60
Group B	85	15	100
Group C	316	50	366
Group D	800	316	1116
Group E	1600	800	2400
Group F	2400	1000	3400
Group G	3567	1000	4567

After training proposed model with 3567 sample and test proposed model with 1000 sample its accuracy showed in next table.

**Table 5: Model accuracy**

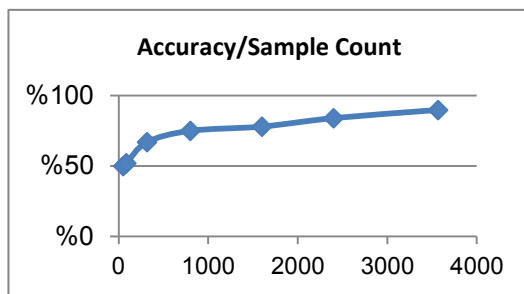
Sample s count	Accurac y	F--measur e	Recal l	Precisio n	Hidde n layers count
1000	89.8%	92.2%	93.8 %	88.9%	60

The experiments prove that there is relation between DNN layers count and the accuracy of proposed model for example when the number layers was 20 the accuracy was 81%. When the layers increased to 100 the accuracy becomes 84% . Finally when the layers fixed in 120 the accuracy reach to 90%. The Next figure shows the relation between hidden layers count and accuracy. It is found the accuracy increases by increasing in number of hidden layers, but it needs more time to learn.



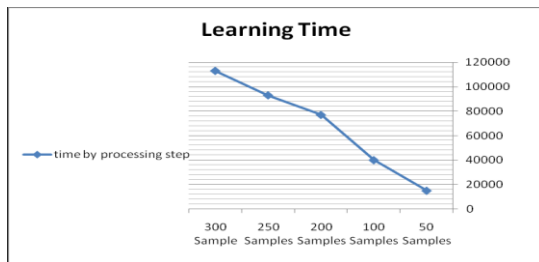
**Fig 5: relation between accuracy and hidden layers count**

The Next figure shows the relation between samples counts and accuracy. Also, it is found the accuracy increases by increasing number of samples, but it needs more time to learn.



**Fig 6: relation between accuracy training samples count**

The Next figure shows the relation between the Time of learning Process and the Samples count.



### 4.3 Results Comparison

**Table 5: Comparison models**

Model name	REV_ID
Combining Axiom Injection and Knowledge Base Completion for Efficient Natural Language Inference	<b>M1</b>
Neural Natural Language Inference Models Enhanced with External Knowledge	<b>M2</b>
A Hybrid Approach to Textual Entailment Recognition	<b>M3</b>
UIO-Lien: Entailment Recognition using Minimal Recursion Semantics	<b>M4</b>
Semantic Parsing for Textual Entailment	<b>M5</b>
Addressing Limited Data for Textual Entailment Across Domains	<b>M6</b>
Proposed model	<b>MY_M</b>

**Table 5: Comparison Results**

Model_REV	Accuracy	Precision	Recall	F-measure
<b>M1</b>	83.45 %	95.75%	64.47%	77.04%
<b>M2</b>	88.6%	--	--	---
<b>M3</b>	75%	75%	75%	75%
<b>M4</b>	77.14%	84.22 %	72.64%	78%
<b>M5</b>	77.1%	81.1%	72.7%	76.6%
<b>M6</b>	61%	62%	48%	54%_
<b>MY_M</b>	89.9%	88.9%	93.8%	92.2%

The previous table shows the comparison between the proposed model and some different other models. Each one of them uses different approach. The first approach [6] is based on Knowledge Base Completion (KBC), Where the Second model [7] is based on external knowledge in co-attention, local inference collection, and inference composition components. The advantage of using external knowledge is more significant when the size of training data is restricted, suggesting that if more knowledge can be obtained, and it may bring more benefit. The third model is based on using Minimal Recursion Semantics (MRS) representations. Minimal Recursion Semantics (MRS) is a framework for computational semantics which can be used for both parsing and generation. A hybrid technique to identify the entailment relation between texts and hypothesis are used in forth model. This technique includes an approach based on lexical similarities and an approach based on the classifier of support vector machine. The approach based on lexical similarities is to use the similarities between a set of words within a text and a set of words within a hypothesis. The previous table shows that the proposed model achieved better accuracy than other models.

### 5. REFERENCES

- [1] Mohamed H. Haggag, Hala Abd Al-Galel, Ahmed M. Ahmed, Semantic Roles Labeling Based on Link Parser, The First International Conference for Computing & Informatics, icci 2012 .
- [2] I.Androutsopoulos , P. Malakasiotis , “A Survey of Paraphrasing and Textual Entailment Methods”, Journal of Artificial Intelligence Research 38 (2010) 135-187 , Submitted 12/09; published 05/10.
- [3] R. Adams. "Textual entailment through extended lexical overlap". In Proceedings of the Second PASCAL Challenges Workshop on Recognising Textual Entailment, pages 128{133, Venice, Italy, April 2006.
- [4] Chu-Carroll, J, and R. Carpenter (1999), Vector-Based Natural Language Call Routing. Journal of Computational Linguistics, 25(30), pp. 361-388, 1999
- [5] J.Bos and K.Markert. “When logical inference helps determining textual entailment” (and when it doesn't). In Proceedings of the Second PASCAL RTE Challenge, Venice, Italy, April 2006.
- [6] Masashi Yoshikawa, Koji Mineshima, ” Combining Axiom Injection and Knowledge Base Completion for Efficient Natural Language Inference”, Association for the Advancement of Artificial Intelligence 2018.

- [7] Qian Chen, Xiaodan Zhu, Neural Natural Language Inference Models Enhanced with External Knowledge, Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Long Papers), pages 2406–2417 Melbourne, Australia, July 15 - 20, 2018
- [8] Chaitanya Shivade, Preethi Raghavany and Siddharth Patwardhan, Addressing Limited Data for Textual Entailment Across Domains, internship at IBM 2016
- [9] Richard Socher, Alex Perelygin, Jean Y. Wu, Jason Chuang, Recursive Deep Models for Semantic Compositionality Over a Sentiment Treebank, Rukmani Ravisundaram and Tayyab Tariq 2013
- [10] Chen Lyu, Yanan Lu, Donghong, Deep Learning for Textual Entailment Recognition, 2015 IEEE 27th International Conference on Tools with Artificial Intelligence.
- [11] Daniel D. Sleator and Davy Temperley, Parsing English Sentence using Link Grammar 2003
- [12] Jorge, G., and Eduardo, M., 2008. "Web-Based Measure of Semantic Relatedness", 9th international conference on Web Information Systems Engineering, Springer-Verlag Berlin, pp. 136-150.
- [13] Li Y., Bandar Z.A., and McLean D. 2003. An Approach for Measuring Semantic Similarity between Words Using Multiple Information Sources. IEEE Trans. on Knowledge and Data Engineering, 15(4), 871-882.
- [14] R. Bar-Haim, J. Berant, and I. Dagan. "A compact forest for scalable inference over entailment and paraphrase rules". In Proceedings of the 2009 Conference on Empirical Methods in Natural Language Processing: Volume 3, pages 1056{1065, Singapore, August 2009. Association for Computational Linguistics.
- [15] M. Sammons, V.G.V. Vydiswaran, T. Vieira, N. Johri, M.W. Chang, D. Goldwasser, V. Srikumar, G. Kundu, Y. Tu, K. Small, et al. Relation "alignment for textual entailment recognition". In Proceedings of the 2009 Text Analysis Conference (TAC'09), Gaithersburg, Maryland, USA, November 2009.
- [16] Elisabeth Lien, Milen Kouylekov, UIO-Lien: Entailment Recognition using Minimal Recursion Semantics, Proceedings of the 8th International Workshop on Semantic Evaluation (SemEval 2014), pages 699–703, Dublin, Ireland, August 23-24, 2014.
- [17] Julio Javier Castillo, A Machine Learning Approach for Recognizing Textual Entailment in Spanish, Proceedings of the NAACL HLT 2010 Young Investigators Workshop on Computational Approaches to Languages of the Americas, pages 62–67, Los Angeles, California, June 2010. c 2010 Association for Computational Linguistics.
- [18] Callison-Burch, C., Osborne, M., and Koehn, P. (2006). Re-evaluation the role of bleu in machine translation research.
- [19] MacCartney, B. (June, 2009). Natural Language Inference, Ph.D. thesis. Stanford University.
- [20] Prodromos Malakasiotis, PARAPHRASE AND TEXTUAL ENTAILMENT RECOGNITION AND GENERATION, Ph.D. thesis, ATHENS UNIVERSITY OF ECONOMICS AND BUSINESS.
- [21] Zuzana Nevěřilová, Paraphrase and Textual Entailment Generation in Czech, Ph.D. thesis, Masaryk University \_ Brno \_ 2014.
- [22] R. Adams. Textual entailment through extended lexical overlap. In Proceedings of the Second PASCAL Challenges Workshop on Recognising Textual Entailment, pages 128{133, Venice, Italy, April 2006.