Hochschule Bonn-Rhein-Sieg

R&D Proposal

Recognizing textual entailment : A comprehensive evaluation of the existing state of the art techniques

Ramit Sharma
Matrikel Number : 9030410

November 26, 2017

First Supervisor : Prof. Dr Paul G. Plöger
Second Supervisor :M.Sc. Deebul Nair

Abstract

Textual entailment is defined as a directional relationship between pairs of text expressions, denoted by T - the entailing Text, and H - the entailed Hypothesis.T entails H if the meaning of H can be inferred from the meaning of T, as would typically be interpreted by people.The Hypothesis H of an entailment pair contradicts the Text T if a human reader would say that the relations/events described by H are highly unlikely to be true given the relations/events described by T. The three-way RTE task requires that systems label each entailment pair as either Entailed,Contradicted, or Unknown.Many natural language processing applications, like Question Answering (QA), Information Extraction (IE), (multi-document) summarization and machine translation (MT) evaluation, need to recognize that a particular target meaning can be inferred from different text variants. Typically entailment is used as part of a larger system, for example in a prediction system to filter out trivial or obvious predictions.The main focus of the work would be to do a survey on the existing state of the art techniques used for recognizing textual entailment and to do a comparative evaluation between any of the two techniques .The work will also look into the application of RTE for short answer grading.

Introduction

- Textual entailment is defined as a directional relationship between pairs of text expressions, denoted by T - the entailing Text, and H - the entailed Hypothesis.
- T entails H if the meaning of H can be inferred from the meaning of T, as would typically be interpreted by people.
- The Hypothesis H of an entailment pair contradicts the Text T if a human reader would say that the relations/events described by H are highly unlikely to be true given the relations/events described by T.
- The three-way RTE task requires that systems label each entailment pair as either Entailed,Contradicted, or Unknown.

Example

Text	Judgments	Hypothesis
A man inspects the uniform of a figure in some East Asian country.	contradiction C C C C C	The man is sleeping
An older and younger man smiling.	neutral N N E N N	Two men are smiling and laughing at the cats playing on the floor.
A black race car starts up in front of a crowd of people.	contradiction C C C C C	A man is driving down a lonely road.
A soccer game with multiple males playing.	entailment E E E E E	Some men are playing a sport.
A smiling costumed woman is holding an umbrella.	neutral N N E C N	A happy woman in a fairy costume holds an umbrella.

Figure 1: Example of Recognizing textual entailment (Image taken from: https://nlp.stanford.edu/)

Related work

- In [1], Yichen Gongyz, Heng Luoz and Jian Zhangz presented Interactive Inference Network (IIN), a novel class of neural network architectures that is able to achieve high-level understanding of the sentence pair by hierarchically extracting semantic features from interaction space. They showed that an interaction tensor (attention weight) contains semantic information to solve natural language inference, and a denser interaction tensor contains richer semantic information. They implemented on instance of such architecture, Densely Interactive Inference Network (DIIN) and demonstrated the state-of-the-art performance on large scale NLI copora and large-scale NLI alike corpus. It's noteworthy that DIIN achieve a greater than 20% error reduction on the challenging Multi-Genre NLI (MultiNLI) dataset with respect to the strongest published system.

- In [2] ,Qian Chen, Xiaodan Zhu, Zhenhua Ling, Si Wei, Hui Jiang and Diana Inkpen presented sequential inference models based on chain LSTMs. They claim that it can outperform all previous models. They further show that by explicitly considering recursive architectures in both local inference modeling and inference composition, they achieve additional improvement. Particularly, incorporating syntactic parsing information contributes to their best result—it further improves the performance even when added to the already very strong model.

- In [3] ,Senlin Zhang, Siyang Liu and Meiqin Liu proposed a long short-term memory (LSTM) model with Sentence Fusion architecture for NLI task.Instead of modifying the internal structure of the LSTM recurrent neural network (RNN) model, they focused on how to make full use of the distributed expression of sentence generated by the LSTM encoder. They improved the performance of basic LSTM recurrent neural networks on Stanford natural language inference (SNLI) corpus by adding Sentence Fusion modules which enrich the distributed expression of sentence generated by the LSTM. Their results demonstrate that the LSTM with Sentence Fusion which reads premise and hypothesis to produce a final fusion representation from which a three-way classifier predicts label has a better performance than LSTM RNN encoders and Lexicalized classifier.

- In [4],Lili Mou, Rui Men, Ge Li, Yan Xu, Lu Zhang, Rui Yan and Zhi Jin propose the TBCNN-pair model to recognize entailment and contradiction between two sentences. In their model, a tree-based convolutional neural network (TBCNN) captures sentence-level semantics; then heuristic matching layers like concatenation, element-wise product/difference combine the information in individual sentences. Experimental results show that their model outperforms existing sentence encoding-based approaches by a large margin.

- In [5] Qian Chen , Xiaodan Zhu , Zhen-Hua Ling , Si Wei , Hui Jiang and Diana Inkpen presented an approach where sentence is represented as a fixed-length vector with neural networks and the quality of the representation is tested with a natural language inference task. Their model is equipped with intra-sentence gated-attention composition which helps achieve a better performance.THey have also tested it on the Stanford Natural Language Inference (SNLI) dataset. They obtain an accuracy of 85.5%, which is the best reported result on SNLI when cross-sentence attention is not allowed.

- In [6] Johan Bos and Katja Markert use logical inference techniques for recognising textual entailment. As the performance of theorem proving turns out to be highly dependent on not readily available background knowledge, they incorporate model building, a technique borrowed from automated reasoning, and show that it is a useful robust method to approximate entailment. Finally, we use machine learning to combine these deep semantic analysis techniques with simple shallow word overlap; the resulting hybrid model achieves high accuracy on the RTE testset, given the state of the art. Our results also show that the different techniques that we employ perform very differently on some of the subsets of the RTE corpus and as a result, it is useful to use the nature of the dataset as a feature.

- In [7] M Kouylekov and B Magnini, estimated the cost of the information of the hypothesis which is missing in the text and can not be matched with entailment rules.They have tested different system settings for calculating the importance of the words of the hypothesis and investigated the possibility of combining them with machine learning algorithm.

- In [8] Stefan Harmeling introduce a system for textual entailment that is based on a probabilistic model of entailment. The model is defined using some calculus of transformations on dependency trees, which is characterized by the fact that derivations in that calculus preserve the truth only with a certain probability. They also describe a possible set of transformations (and with it implicitly a calculus) that was successfully applied to the RTE3 chal-lenge data. However, our system can be improved in many ways and they see it as the starting point for a promising new approach to textual entailment.

- In [9] Jihun Choi, Kang Min Yoo and Sang-goo Lee assert that recursive neural networks (RvNNs) have been shown to be suitable for representing text into fixed-length vectors and achieved good performance on several natural language processing tasks. However, the main drawback of RvNNs is that they require structured input, which makes data preparation and model implementation hard. In this paper, they propose Gumbel Tree-LSTM, a novel tree-structured long short-term memory architecture that efficiently learns how to compose task-specific tree structures only from plain text data. Our model uses Straight-Through Gumbel-Softmax estimator to decide the parent node among candidates dynamically and to calculate gradients of the discrete decision. They evaluate the proposed model on natural language inference and sentiment analysis, and show that their model outperforms or is at least comparable to previous models. They also found that their model converges significantly faster than other models.

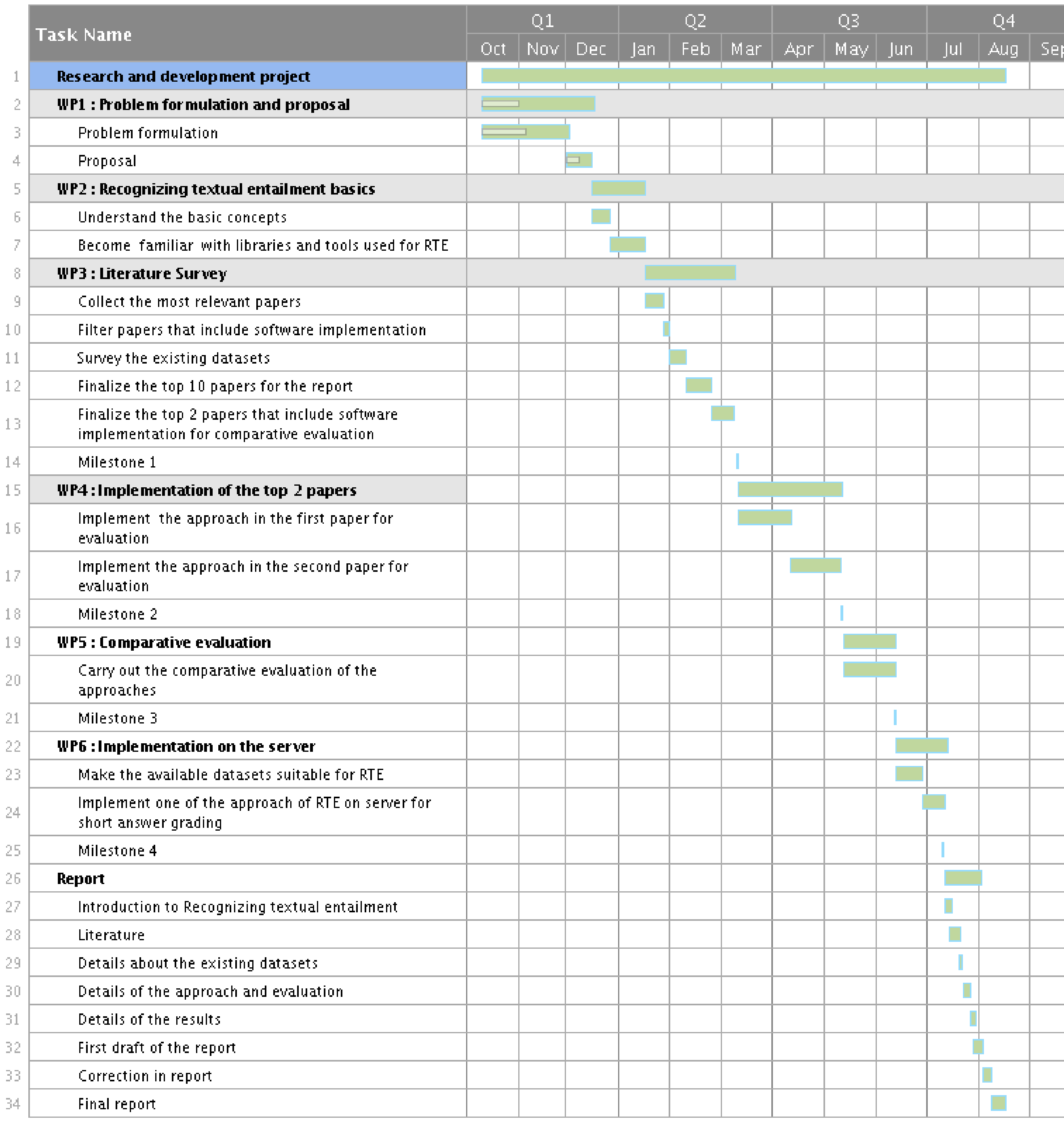
Problem Statement

- The project is focused on comprehensive evaluation of the existing state of the art techniques for recognizing textual entailment
- The first step would be to do a survey on the existing state of the art techniques for recognizing textual entailment
- The next step would be to go through the datasets available for recognizing textual entailment
- Then the third step would be to implement two approaches and do a comparative evaluation of the methods.
- This project also intends to look for the application of the textual entailment for short answer grading

Expected Goals

- Minimum
  - To carry out the survey of the existing state of the art techniques for recognizing textual entailment
  - To carry out the survey for the existing datasets used for recognizing textual entailment
- Expected
  - To implement any of the two techniques used for recognizing textual entailment
  - To do the comparative evaluation of the two techniques which are implemented
- Maximum
  - Implement the approach for short answer grading in a development server

Work plan



Task Name
Q1
Q2
Q3
Q4
Oct
Nov
Dec
Jan
Feb
Mar
Apr
May
Jun
Jul
Aug
Sep
1 Research and development project
2 WP1 : Problem formulation and proposal
3 Problem formulation
4 Proposal
5 WP2 : Recognizing textual entailment basics
6 Understand the basic concepts
7 Become familiar with libraries and tools used for RTE
8 WP3 : Literature Survey
9 Collect the most relevant papers
10 Filter papers that include software implementation
11 Survey the existing datasets
12 Finalize the top 10 papers for the report
13 Finalize the top 2 papers that include software implementation for comparative evaluation
14 Milestone 1
15 WP4 : Implementation of the top 2 papers
16 Implement the approach in the first paper for evaluation
17 Implement the approach in the second paper for evaluation
18 Milestone 2
19 WP5 : Comparative evaluation
20 Carry out the comparative evaluation of the approaches
21 Milestone 3
22 WP6 : Implementation on the server
23 Make the available datasets suitable for RTE
24 Implement one of the approach of RTE on server for short answer grading
25 Milestone 4
26 Report
27 Introduction to Recognizing textual entailment
28 Literature
29 Details about the existing datasets
30 Details of the approach and evaluation
31 Details of the results
32 First draft of the report
33 Correction in report
34 Final report

References

[1] Yichen Gongyz, Heng Luoz, Jian Zhangz. *"Natural Language Inference Over Interaction Space"*. Cornell University Library,13 Sep 2017

[2] Qian Chen, Xiaodan Zhu, Zhenhua Ling, Si Wei, Hui Jiang, Diana Inkpen. *"Enhanced LSTM for Natural Language Inference"*. Cornell University Library,20 Sep 2016

[3] Senlin Zhang, Siyang Liu, Meiqin Liu. *"Natural language inference using LSTM model with sentence fusion"*. IEEE ,11 September 2017

[4] Lili Mou, Rui Men, Ge Li, Yan Xu, Lu Zhang, Rui Yan, Zhi Jin. *"Natural Language Inference by Tree-Based Convolution and Heuristic Matching"*. Cornell University Library ,13 May 2016

[5] Qian Chen , Xiaodan Zhu , Zhen-Hua Ling , Si Wei , Hui Jiang , Diana Inkpen . *"Recurrent Neural Network-Based Sentence Encoder with Gated Attention for Natural Language Inference"*. Association for Computational Linguistics ,September 2017

[6] Johan Bos,Katja Markert. *"Recognising Textual Entailment with Logical Inference"*. HLT '05 Proceedings of the conference on Human Language Technology and Empirical Methods in Natural Language Processing ,October 2005

[7] M Kouylekov, B Magnini. *"Tree Edit Distance for Recognizing Textual Entailment: Estimating the Cost of Insertion"*. Proc. of the PASCAL RTE-2 Challenge, 2006

[8] Stefan Harmeling. *"An extensible probabilistic transformation-based approach to the third recognizing textual entailment challenge"*. Proceeding RTE '07 Proceedings of the ACL-PASCAL Workshop on Textual Entailment and Paraphrasing

[9] Jihun Choi, Kang Min Yoo, Sang-goo Lee. *"Learning to Compose Task-Specific Tree Structures"*. Cornell University Library ,10 Jul 2017

[10] $https://en.wikipedia.org/wiki/Textual\_entailment.$

[11] $https://tac.nist.gov//2011/RTE/.$

[12] $https://nlp.stanford.edu/projects/snli/$