

SUBMISSION OF WRITTEN WORK

Class code: KSTPSDT1KU
 Name of course: Thesis Preparation SDT
 Course manager: Marco Carbone
 Course e-portfolio:

Thesis or project title:
 Supervisor:

Full Name:
Magnus Stahl Jacobsen

Birthdate (dd/mm-yyyy):
26/02-1991

E-mail:
msta @itu.dk

- | | | |
|----------|-------|---------------|
| 1. _____ | _____ | _____ @itu.dk |
| 2. _____ | _____ | _____ @itu.dk |
| 3. _____ | _____ | _____ @itu.dk |
| 4. _____ | _____ | _____ @itu.dk |
| 5. _____ | _____ | _____ @itu.dk |
| 6. _____ | _____ | _____ @itu.dk |
| 7. _____ | _____ | _____ @itu.dk |

Thesis Preparation Report

Magnus Stahl (msta@itu.dk)

IT-University of Copenhagen

Fall 2016

Supervisor: Natalie Schluter

Contents

1	Introduction	3
1.1	Research questions	4
1.2	Motivation	5
2	Work so far	5
2.1	Preparation	5
2.2	Work	6
3	Thesis plan	7
3.1	Initial research	7
3.2	Baseline implementation and method	7
3.3	Experimentation	8
3.4	Write up, report	8
4	Conclusion	8
	List of Figures	9

1 Introduction

This report details introductory thesis work as part of a spring 2017 thesis at the IT-University of Copenhagen. The report will describe the work undertaken so far, and a plan for how the thesis will be completed in the spring.

The thesis I will write is placed in the field of Natural Language Processing, and more specifically Natural Language Understanding. Natural Language Understanding is the subject of making computers understand human language, possibly on the same level as humans. The subject can be investigated in a strong or weak AI perspective. This either entails trying to create a general idea of intelligence (strong AI) or splitting the language understanding task into smaller subtasks which can then be solved. My thesis will be placed in the latter category. I will look only at sentence level language understanding. Understanding a sentence is a task also composed of many other subtasks, from spelling individual words and understanding the meaning, recognizing composite words, understanding how each word relate to the sentence and abstracting higher meanings from the sentence. I will look at the task of *classifying* relations between entities (usually nouns) in a sentence. Classification of relations in sentences are important because relations are a central component in the higher meaning for a sentence. Classification in this context means mapping sentences and their entities to a category which explains the nature of the relationship between them. Classification is most commonly solved with machine learning, which in this case entails looking at a set of sentences with known relations, and then extracting patterns automatically from these samples.

The task of relation classification can be more formally defined as: Given a sentence S and two labeled entities $e1$ and $e2$, classify the relation between the entities and (optionally) the ordering of the relation. Here is an example:

S : “*The train rolled onto the platform with a peculiar choo choo.*”

$\langle e1 \rangle$: *train*

$\langle e2 \rangle$: *platform*

Relation: *Entity-Destination($e1, e2$)*

Many classes of algorithms exists for solving classification tasks, but in my thesis I will focus on neural networks. Artificial Neural Networks (ANN) are a group of data structures and related algorithms that loosely models features of a biological brain. ANNs can be used for a wide variety of machine learning tasks such as text analysis, speech recognition and computer vision. The last 10 years have witnessed a great increase in the use of neural networks due to multiple factors such as increased computing power and parallel capabilities with mainstream GPU's. Another factor is the

work of specific research groups on new neural network patterns such as Long-Short-Term-Memory or Convolutional Neural Networks by the IDSIA group [7].

In my thesis, I will investigate contemporary neural networks used for solving the relation classification task. I will try to recreate some of the results claimed by the creators of those networks by re-implementing the networks with modern deep learning frameworks. Then I will investigate the state-of-the-art approaches and experimentally find out if the same results may be achieved with a simpler network. I will try to investigate why certain approaches are suggested and try to separate parts of the network structure which is badly (or perhaps un-)motivated. I will use a common dataset (SemEval 2010, task 8) to document the results. And finally, if time allows it, I will perhaps look at relation classification in a domain-based space where the relations to be classified are very specific to the individual domain.

1.1 Research questions

I have formulated the following tentative research questions:

- What are the current state-of-the-art systems for relation classification with neural networks?
- Is it possible to simplify structures in state-of-the-art systems while maintaining (or increasing) the F1 metric?

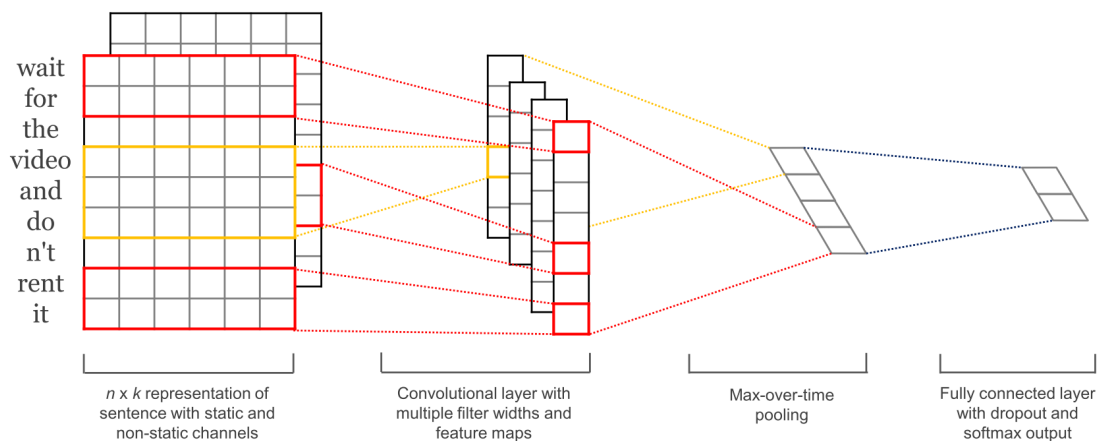


Fig. 1: Relation Classification with Neural Networks

1.2 Motivation

This section motivates why the subject of the thesis is important.

Relation classification is a hard task for computers. Correctly classifying entities in language requires understanding the context of the entities as well as an abstract understanding of their relationship. Recent deep learning experiments have yielded excellent results compared to feature-based classifiers, which require feature engineering. Neural networks are interesting in NLP because they allow features to be learned automatically from raw input data. This can lower the entry barrier for working with language systems as features usually require linguistic domain knowledge.

Simplification of state-of-the-art systems, and especially neural networks, is interesting because the newest systems are often built on top of other state-of-the-art systems. These “research iterations” often lead to including substructures from previous systems which might not be relevant in the newer system. Additionally, hyperparameters are not always decided by search but rather by sticking to values which were chosen in a previous system. Hyperparameters might incur unnecessary memory or performance costs. Finally, many papers suggest that simpler networks reduce overfitting and thus generalize better [8].

2 Work so far

This section details how I have been preparing for the thesis until January 2017. I will begin by detailing the literature I have followed to learn more about the subject.

2.1 Preparation

I have before the start of the 2016 fall semester completed courses in mathematics (msc level), data mining and intelligent systems. I have written a bachelors thesis on the subject of classification and prediction.

In the thesis preparation course, I have learned more about deep learning and NLP. I have completed a small deep learning course on Udacity [2]. To complement the general deep learning knowledge I have started to read an MIT newly released grad level book about deep learning [1]

To connect my knowledge about machine learning with natural language processing, I started with Yoav Goldberg’s tutorial about Deep Learning in NLP (reference). I have followed a coursera course on Natural Language Processing from Columbia University which details the more traditional approaches to NLP tasks. I have attended a NLP workshop on structured prediction in Named Entity Recognition and Part-Of-Speech tagging.

Finally I have researched the algorithms behind word embeddings which are a commonly used tool when transforming text input to real-valued vectors. The word embedding algorithms which I have investigated (and which are also used in the baseline) are the skip-gram and the word2vec model [4].

2.2 Work

Now I will describe the work I have done on the first research question described in section 1.

I have researched which research papers claims state-of-the-art performance on the relation classification task. Most papers use a convolutional structure as the base structure of the network. Several papers in the last two years have claimed a new F1 score record on the SemEval 2010 task 8 dataset. I have implemented a few of those [10] [6] and read others who have similar scores with slightly different approaches (other papers here). The papers might differ in their use of loss function, input representation, values of hyperparameters and training algorithms.

I have also read a paper detailing the use of LSTM structures as their base structure [5]. Some papers differ slightly in their objective (another common objective is relation extraction, in which the entities are not known as part of the input).

I have implemented these novel networks in Keras, a deep learning framework which uses TensorFlow as backend [9] [3].

3 Thesis plan

This sections details my plan for the thesis. During the entire project I plan to have weekly meetings with my supervisor. Before starting the thesis, I will also revise the project agreement so it encompasses the experiences I have gathered during the thesis preparation course.

I will now briefly outline the overall plan and then go into details with each step. Shown below is a Gantt chart which details the overall plan:

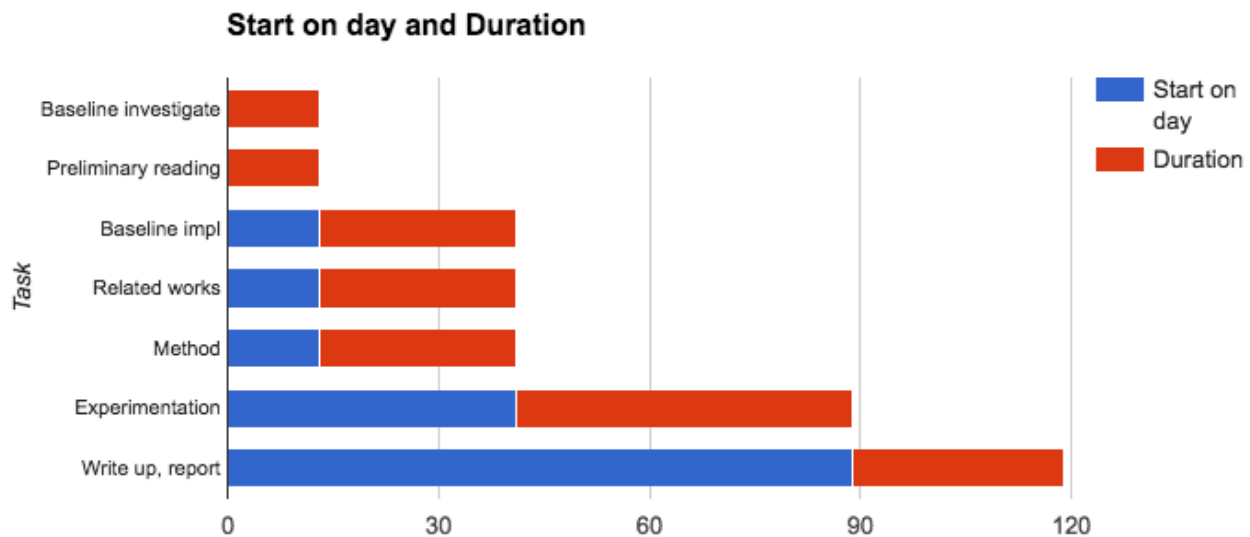


Fig. 2: Gantt Chart for Thesis

As shown on the chart, I have divided my work in 7 parts, which will be done in four timespans.

3.1 Initial research

The first few weeks of the project will be used on additional preliminary reading and finishing the investigation on current state-of-the-arts systems. The conclusion of the investigation will be an outline of a system that will serve as my baseline for future experimentation and simplification.

3.2 Baseline implementation and method

The next part of the project is designated to be roughly one month in length. The primary goal is to write the part of the report that details the related works (which I know about from the baseline

investigation) and implement the baseline implementation. The end product (except from the report) of this part is a working baseline that is on-par with the current state-of-the-art systems. In this part I will also work on a method for simplifying the existing implementations.

3.3 Experimentation

The third timespan of the project will be spent on simplifying current state-of-the-art systems. I will investigate different substructures of the proposed networks in terms of their motivation and impact on the result. For the experimentation I will use a server which is provided by the NLP group at ITU. The server provides the necessary GPU to efficiently parallelize neural network computations.

3.4 Write up, report

The final part of the project will be focused on writing a good report. While I will write the report during the rest of the project as well, a final month of focusing on detailing everything correctly seems necessary. In this period I will also have time to properly document and illustrate the (potentially) new structure I have created. Finally, I am reserving the last week of the project for proof-reads.

4 Conclusion

This section concludes my report. I have detailed the objective of the thesis and motivated why it is important. I have placed the subject of the thesis in the area of science where it belongs. I have formulated tentative research questions and defined the task which I'm trying to solve formally. I have detailed the preparatory work I have done as part of the course, and finally I have detailed the plan of the full thesis project.

References

- [1] Ian Goodfellow, Yoshua Bengio, and Aaron Courville. *Deep Learning*. MIT Press, 2016.
<http://www.deeplearningbook.org>.
- [2] Google. Deep learning course by google, 2016. <https://www.udacity.com/course/deep-learning--ud730>.
- [3] "Google Inc.". Tensorflow: An open-source software library for machine intelligence, 2016.
<https://www.tensorflow.org/>.

- [4] Tomas Mikolov, Kai Chen, Greg Corrado, and Jeffrey Dean. Efficient estimation of word representations in vector space, 2013. <https://arxiv.org/pdf/1301.3781.pdf>.
- [5] Makoto Miwa and Mohit Bansal. End-to-end relation extraction using lstms on sequences and tree structures, 2016.
- [6] Thien Huu Nguyen and Ralph Grishman. Relation extraction: Perspective from convolutional neural networks, 2015.
- [7] Jrgen Schmidhuber. Recurrent neural networks. <http://people.idsia.ch/~juergen/rnn.html>. Accessed: 04-01-2017.
- [8] Nitish Srivastava, Geoffrey Hinton, Alex Krizhevsky, Ilya Sutskever, and Ruslan Salakhutdinov. Dropout: A simple way to prevent neural networks from overfitting, 2014.
- [9] Keras Dev Team. Keras: Deep learning library for theano and tensorflow, 2016. <https://keras.io/>.
- [10] Linlin Wang, Zhu Cao, Gerard de Melo, and Zhiyuan Liu. Relation classification via multi-level attention cnns, 2016.

List of Figures

Fig. 1 (p. 4): Relation Classification with Neural Networks. From: Kim, Yoon. "Convolutional Neural Networks for Sentence Classification, 2014"

Fig. 2 (p. 7): Gantt Chart for Thesis. From: Magnus Stahl