Summer_Paper_Reading_2016

eg: 0-ACL16-Ng-Paper_title.pdf

- Abstract
  - Overview:
  - Advantage:
  - Disadvantage:
  - What can I do? / Can I employ its idea?
- Experiments
  - DataSet:
  - Toolkit:
  - Baseline:
  - Result:

1-NIPS15-Google-Pointer Network

PDF, Bib

- [**Problem-Paper**] the number of target classes depends on the length of input, which is variable.
- [**Problem-Experiment**]
  - finding planar convex hulls
  - computing Delaunary triangulations
  - the planar Travelling Salesman Problem
- [**Model**] instead of using attention to blend hidden units of an encoder to a context vector of a decoder, it uses attention as a pointer to select a member of the input sequence as the output.
- [**Remark**]
  - Neural attention model deals with the fundamental problem representing variable length dictionaries, appling to three distinct algorithmic problems.
  - Idea is fantastic.

Authors of this paper propose an attention model based on "Neural machine translation by jointly learning to align and translate"[1] by using LSTMs. Their model decides what to predict by using the output of a softmax MLP over the inputs as opposed to regular attention model used in [1](RNNSearch), where they used that to use for the convex combination of the input annotations to predict where to attend to translate in the target language. The output of the softmax-MLP predicts the location of the output at each timestep directly from the input sequence.

The advantage of this approach is that, the output softmax dictionary only depends on the length of the input sequence instead of on the number of possible instances in the problem. The authors have provided very interesting experimental results on three different discrete optimization problems.

2-ACL16-IBM-Addressing Limited Data for Textual Entailment Across Domains

PDF, !Bib

- [**Problem-Paper**] exploit unlabled data to improve F-score for TE task.
- [**Problem-Experiment**] find all sentences in a corpus that entail a given hypothsis.
  Domain: Newswire (RTE-5,6,7) & Clinical (self construct)
- [**Model**] Tradition Features + self-training/activate learning
- [**Conclusion**] The author analysis the possible explanation for the improvement: Class Distribution. down-sampling and up-sampling is not useful. Activate learning will sampling positive examples, thus it can match the performance with fewer examples.
- [**Remark**] Experiment is beautiful and convising, Information Retrival Method.

This paper illustrates that how self-training will influence the classification and why active learning will reduce the examples to only 6 percent.
The author experiments on two domains -- newswire and clinical. First, the author creates an entailment dataset for clinical domain with human annoted. Second, he builds a highly competitive supervised system, called ENT. Last, he explore two strategies - self-training and active learning to address the lack of labled data. Experiment is done in detail and convincing.

3-ACL16-Stanford-A Thorough Examination of the CNN/Daily Mail Reading Comprehension Task

PDF, !Bib, !Github

- [**Problem**] CNN/Daily Mail Reading Comprehension Task
- [**Model**]
    - Traditional Features
        - Feature ablation analysis
    - Attention Neural Network(followed 5-NIPS15-NYU-End-To-End Memory Networks)
- [**Related Dataset**]
    - CNN/Daily Mail (2015)
    - MCTest（2013）
    - Children Book Test （2016）
    - bAbI （2016）
- [**Data Analysis**] breakdown of the Examples, Analysis the perfomance on each categories(although on small dataset).
- [**Remark**] Also we can construct traditional ML and NN, **data analysis is important**, without this, Experiment seems to be inconvincing.

This paper conducts a thorougn examination of CNN/Daily Mail Reading Comprehension Task, which origin from the idea that a bullet point usually summaries one or several aspects of the article. **If the computer understands the content of the artticle. It should be able to infer the missing entity in the bullet point.**

two supervised systems are implemented -- a conventional entity centric classfier and an end to end neural network. Expriment shows that the straight-forward NLP system, compared with origin frame-semantic parser[^1], obtain accuracies of 72.4% and 75.8％ on these two dataset ｓ.

Besides, the author **extracts 100 examples to analysis the results**. She roughly classify the examples into 6 categories, i.e., Exact Match, Paraphrase, Parial clue, Multiple sent, (Coref.Error, Hard), the last two is hard for human to obtain the correct answer.

[^1]: Teaching Machine to read and comprehend, NIPS15, Hermann et.al

4-NAACL16-CMU-Hierarchical Attention Networks for Document Classification

PDF, Bib

5-NIPS15-NYU-End-To-End Memory Networks

PDF, Bib, Theano, Tensorflow

- [**Goal**] introdce a **neural network** with a **recurrent attention** over a possibly large ***external memory***.
- [**Problem-Experiment**]
    - a). synthetic question answering (Reading Comprehension Task)
    - b). Language Model
- [**Dataset**]
    - a).bAbI
    - b).Penn Tree Bank & Text8
- [**Model**]
  Notation:
    - input: sentences: $$x*1, x*2, \ldot, x_i$$, question: q
    - output: answer:a
    - Variable: A, B, C, W
    - $$ shape(x*i) = d; shape(A) = (d, V); shape(m*i) = d; shape(input) = (len(sentences), d) $$
    - Sentence Representation: BoW & **Position Encoding** (slightly different from BoW, add position information), $$ m_i = \sum*{j}Ax*{ij} $$
    - Temporal Encoding: QA tasks require some notion of temporal context, $$ m*i = \sum*{j}Ax*{ij} + T*{A}(i)$$
    - Random Noise: to reqularize T_A, randomly add 10% of empty memories to the stories.
    - The capacity of memory is restricted to the most recent 50 sentences.
    - **Since the number of sentences and the number of words per sentence varied between problems, a null symbol(all zero) was used to pad them all to a fixed size**
- [**Remarks**]
    - **How to write a new Model with not the state-of-art performance**?
        - incuction previous model to this model (LSTM, Attention, ...)
        - Compare with other related model (where is the difference?)
        - How can the model changes?
        - What can the model apply?

- **Related Works** deserves to be learned.
- **Code** deserves to be implenment by myself.

This Paper introduce a neural nerwork with a recurrent attention model over a possibly large external memory.

The memory in RNNs model is the state of the network, which is latent and inherently unstable over long timescales. The LSTM-based Model address this through local memory cells which lock in the network from the past. This model differs from these in that it uses a global memory, with shared read and write functions.

This model also related to attention mechanism in Banhadnau's work[^1], although Banhadnau's work[^1] is only over a single sentence rather than many.

This approach is compettive with Memory Networks, but with less supervision.

Next is quoted from yanranli's blog. The summary deserves me to leaned. you can refer to here for more details.

> And the authors attempt several ways in this paper to fulfill their goal. First, the single-layer or multi-layer, and then the transformation of feature space. If one separate the output of the end-to-end memory networks, they can be parallized with typical RNN. The output comprises of two parts, internal ouptut and external output, which can be parallized to RNN's memory and predicted label, respectively.

[^1]: Neural machine translation by jointly learning to align and translate. ICLR15, Bahdanau, Cho, and Bengio

TD1-ACL16-Microsoft-Deep Reinforcement Learning with a Natural Language Action Space

PDF, !Bib

TD2-NIPS15-DeepMind-Teaching Machines to Read and Comprehend

PDF, Bib, Tensorflow

Other Papers:

1-ACL16-Simple PPDB: A Paraphrase Database for Simplification

PDF, !Bib

2-harvard-Visual Analysis of Hidden State Dynamics in Recurrent Neural Networks

PDF, code

LSTM Visual Analysis

NAACL

A Neural Network-Based Abstract Generation Method to Opinion Summarization

Bayesian Supervised Domain Adaptation for Short Text Similarity

Clustering Paraphrases by Word Sense

Convolutional Neural Networks vs. Convolution Kernels: Feature Engineering for Question Answering

DAG-structured Recurent Neural Networks for Semantic Compositionality

Deep LSTM based Feature Mapping for Query Classification

Dependency Based Embeddings for Sentence Classification Tasks

Dependency Sensitive Convolutional Neural Networks for Modeling Sentences and Documents

~~Hierarchical Attention Networks for Document Classification~~

Learning Distributed Representations of Sentences from Unlabelled Data