Neural Architectures for Named Entity Recognition

Guillaume Lample (2016)

Previous Model	Our Model
hand-crafted features	features made by neural architectures
language-specific knolwedge or resources	**No** such knowledge or resources
only supervised learning	both supervised and unsupervised leearning

- Cons
    1. They present a new neural architecture, named Bi-LSTM-CRF, which use both supervised and unsupervised, and do not require domain-specific knowledge and resources.
    2. Word representations (word embedding + character-level) to capture morphological and orthographic information.
    3. This neural architecture provide the best NER results (compared with models that use external resources).
- Pros
    1. No error analysis and speed comparison.
    2. It's not completely end to end, since it still need feature engineering.
    3. Didn't outperform the previous best result in English dataset (90.04, 91.2). Compared to the work of Xuezhe Ma, which achieved F1 score of 91.21 on the CoNLL 2003 English dataset and also achieved end-to-end.

- Future work

    A lot of current work is being done on transfer learning for NER since there are some domains with a lot of annotated data available and others with hardly any at all.

Neural Machine Translation by Jointly Learning to Align and Translate

Dzmitry Bahdanau, Kyunghyun Cho, Yoshua Bengio (ICLR 2015)

- Cons
    1. The encoder is a bidirectional RNN, in which they take the annotation of each word to be the concatenation of the forward and backward RNN states. The idea is that the hidden state should encode information from both the previous and following words.
    2. The proposed attention mechanism is a weighted sum of the input hidden states, the weights for which come from an attention function and are softmax-normalized.
    3. Incorporating the attention mechanism shows large improvements on longer sentences.

4. A big advantage of attention is that it gives us the ability to interpret and visualize what the model is doing.

- Pros
    1. Their model formulation to capture long-term dependencies is far more principled than Sutskever's inverting the input idea. They should have done a comparative study with their approach as well though.
    2. If we do character-level computations and deal with sequences consisting of hundreds of tokens the above attention mechanisms can become prohibitively expensive. It seems like a waste and not at all what human are doing.
    3. This translation benchmarks should also be done in Japanese compared to the French which is similar with English.
- Future work
    1. Handle unknown, or rare words: Copy mechanism
    2. Repetition: Coverage mechanism
    3. An alternative approach to attention is to use Reinforcement Learning to predict an approximate location to focus to.

Globally Coherent Text Generation with Neural Checklist Models

Chlo´e Kiddon (EMNLP 2016)

- Cons
    1. Generates a natural language description for achieving a goal, such as generating a recipe for a particular dish.
    2. Predicts which agenda item is being referred to and stores those predictions for use during generation.
    3. Their adapted GRU has a change to the computation of the new content as follows:

$$\begin{aligned}\tilde{\mathbf{h}}_t &= \tanh(W_h x_t + \mathbf{r}_t \odot U_h \mathbf{h}_{t-1} \\ &\quad + \mathbf{s}_t \odot Y\mathbf{g} + \mathbf{q}_t \odot (\mathbf{1}_L^T Z E_t^{new})^T,\end{aligned}$$

where $\mathbf{s}_t$ is a *goal select* gate and $\mathbf{q}_t$ is a *item select* gate, respectively defined as

$$\mathbf{s}_t = \sigma(W_s \mathbf{x}_t + U_s \mathbf{h}_{t-1}),$$
$$\mathbf{q}_t = \sigma(W_q \mathbf{x}_t + U_q \mathbf{h}_{t-1}).$$

The goal select gate controls when the goal should be taken into account during generation. The item select gate controls when the available agenda items should be taken into account.

- Pros
    1. During training, the author mentioned the weak heuristic supervision on latent variables. But this signal is totally different between the recipe generation task and dialogue generation task. This should be written down in detail.

2. In human evaluation, both the Attention and EncDec baselines and the Checklist model beat the true recipes in terms of having better grammar.
3. The Checklist model generates catch-all phrases like ‘all ingredients’ used in 13% of the generated recipes, whereas only 7.8% of true recipes use that construction.

- Future work

1. Generate more precise referring expressions. Future work could better model the full set of steps needed to achieve the overall goal.
2. The neural checklist model is sensitive to hyper-parameter initialization. It should be investigated in future work.
3. Incorporate referring expressions for sets or compositions of agenda items.

Towards Constructing Sports News from Live Text Commentary

Jianmin Zhang (ACL 2016)

- Cons
    1. A probabilistic sentence selection algorithm is proposed to solve the redundancy problem.
    2. Based on the LTR model using traditional features and Task-specific features.
    3. Formulate the task in a supervised learning to rank framework, utilizing both traditional sentence features for generic document summarization and designed task-specific features.
- Pros
    1. The distinction between different levels is treated consistently in pair-wise. The size of the associated document set brings the model bias.
    2. There are a lot of short sentences in the live text, and the methods taken by the paper are easy to ignore these short sentences, and some short sentences containing key events should be extracted.
    3. The author thinks that the sentence selection algorithm mentioned in this paper can only solve the local redundancy problem and cannot solve the global redundancy problem. But I personally think that the final summary of the news and the process of describing the content of the process is repeatable is acceptable.
- Future work

  There is a room for improvement in readability.