ICLR 2017 Highlights

Natural Language Processing

Maha ELBAYAD

June 2, 2017

Overview

When: April 24 - 26, 2017

Where: Toulon, France

What: Relevant topics

- Representation learning
- Reinforcement learning
- Metric & kernel learning
- Applications in CV; NLP, Robotics,...

Statistics

Overall: A total of 491 papers were submitted and the decisions were:

- Oral: 15 (3%)
- Poster: 183 (37.3%)
- Suggested for workshop: 48 (9.8%)
- Rejected: 245 (49.9%).

Top rated papers:

- (9.67) C. Zhang et al. "Understanding deep learning requires rethinking generalization." (Google)
- (9.0) B. Zoph et al. "Neural architecture search with reinforcement learning." (Google)
- (8.75) M. Arjovsky et al. "Towards principled methods for training generative adversarial networks." (Facebook)

sources: prlz77.github.io, medium.com/@karpathy

Selected papers

- Zhilin Yang et al. "Words or Characters? Fine-grained Gating for Reading Comprehension." (Carnegie Mellon)
- Zhouhan Lin et al. "A structured self-attentive sentence embedding" (MILA & IBM)
- David Krueger et al. "Zoneout: Regularizing RNNS by randomly preserving hidden activations." (MILA)

Words or Characters? Fine-grained Gating for Reading Comprehension

Yang, Z., Dhingra, B., Yuan, Y., Hu, J., Cohen, W. W., & Salakhutdinov, R. (2016)

Motivation:

To combine word-level and character-level presentations, use a fine-grained gating mechanism for a dynamic combination.

Word-level	Character-level
Obtained from a lookup table Each unique token is represented as a vector	Applying RNN or CNN on the characters sequence The hidden states combined to form the representation
Good at memorizing the semantics Requires large amount of data to learn similarities	Suitable for modeling sub-word morphologies Capture the similarities by design Ease of modeling out of vocabuary tokens.

Words or Characters? Fine-grained Gating for Reading Comprehension

Yang, Z., Dhingra, B., Yuan, Y., Hu, J., Cohen, W. W., & Salakhutdinov, R. (2016)

Reading comprehension setting:

Given a document $P = (p_1, p_2, ..p_M)$ and a qeury $Q = (q_1, ...q_M)$ we want to get the index or span of indices in the document that matched the query Q.
A token p_i is denoted as (w_i, C_i) where:

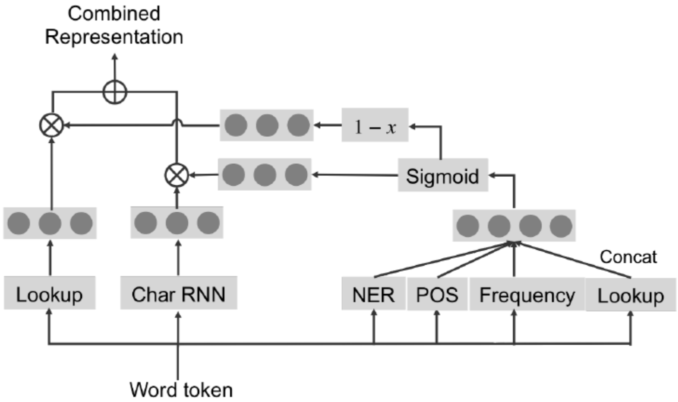
- $w_i \equiv$ the word index in the vocabulary
- $C_i \equiv$ the characters indices.

Besides, for each word w, we feed to the model a vector v encoding its properties e.g. concat(named entity tags, part-of-speech tags, binned document frequency vectors, word-level representation)

Words or Characters? Fine-grained Gating for Reading Comprehension

Yang, Z., Dhingra, B., Yuan, Y., Hu, J., Cohen, W. W., & Salakhutdinov, R. (2016)

Approach:



$$
\begin{aligned}
c &= \text{Encode}(C) \\
Ew &= \text{Encode}(w) \\
g &= \sigma(W_g v + b_g) \\
h &= f(c, w) = g \odot c \\
&\quad + (1 - g) \odot Ew
\end{aligned}
$$

Words or Characters? Fine-grained Gating for Reading Comprehension

Yang, Z., Dhingra, B., Yuan, Y., Hu, J., Cohen, W. W., & Salakhutdinov, R. (2016)

Performance

On the Twitter dataset:

Model	Precision@1	Recall@10	Mean Rank
word (Dhingra et al., 2016b)	0.241	0.428	133
char (Dhingra et al., 2016b)	0.284	0.485	104
word char concat	0.2961	0.4959	105.8
word char feat concat	0.2951	0.4974	106.2
scalar gate	0.2974	0.4982	104.2
fine-grained gate	**0.3069**	**0.5119**	**101.5**

A structured self-attentive sentence embedding

Lin, Z., Feng, M., Santos, C. N. D., Yu, M., Xiang, B., Zhou, B., Bengio, Y. (2017)

Motivation:

Extract an interpretable sentence embedding in 2D with each row attending to a part of the sentence.

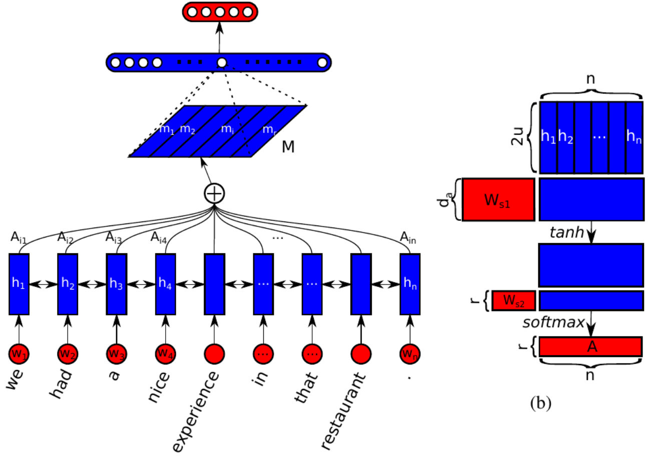
Existing approaches are either:

- Universal sentence embeddings: SkipThought, ParagraphVector,..
- Task specific: usually the final hidden state of an RNN

A structured self-attentive sentence embedding

Lin, Z., Feng, M., Santos, C. N. D., Yu, M., Xiang, B., Zhou, B., Bengio, Y. (2017)

Approach:



Model for sentiment analysis.

M ca suffer from redundancy if the model yields the same summation weights in A.

Regularization:

1) Maximize KLD between any two rows of A. (unstable)

2) Minimize $P = \|AA^T - I\|_F^2$

A structured self-attentive sentence embedding

Lin, Z., Feng, M., Santos, C. N. D., Yu, M., Xiang, B., Zhou, B., Bengio, Y. (2017)

Experiments:

Sentiment analysis - Yelp:
Take the review as input and predict the number of stars associated.
Author profiling - Age:
Input a tweet and predict the age range of the author.

Models	Yelp	Age
BiLSTM + Max Pooling + MLP	61.99%	77.40%
CNN + Max Pooling + MLP	62.05%	78.15%
Our Model	**64.21%**	**80.45%**

A structured self-attentive sentence embedding

Lin, Z., Feng, M., Santos, C. N. D., Yu, M., Xiang, B., Zhou, B., Bengio, Y. (2017)

Experiments: Sentiment analysis

- if I can give this restaurant a 0 I will we be just ask our waitress leave because someone with a reservation be wait for our table my father and father-in-law be still finish up their coffee and we have not yet finish our dessert I have never be so humiliated do not go to this restaurant their food be mediocre at best if you want excellent Italian in a small intimate restaurant go to dish on the South Side I will not be go back
- this place suck the food be gross and taste like grease I will never go here again ever sure the entrance look cool and the waiter can be very nice but the food simply be gross taste like cheap 99cent food do not go here the food shot out of me quick then it go in
- everything be pre cook and dry its crazy most Filipino people be used to very cheap ingredient and they do not know quality the food be disgusting I have eat at least 20 different Filipino family home this not even mediocre
- seriously f *** this place disgust food and shitty service ambience be great if you like dine in a hot cellar engulf in stagnate air truly it be over rate over price and they just under deliver forget try order a drink here it will take forever get and when it finally do arrive you will be ready pass out from heat exhaustion and lack of oxygen how be that a head change you do not even have pay for it I will not disgust you with the detailed review of everything I have try here but make it simple it all suck and after you get the bill you will be walk out with a sore ass save your money and spare your self the disappointment
- i be so angry about my horrible experience at Medusa today my previous visit be amaze 5/5 however my go to out of town and I land an appointment with Stephanie I go in with a picture of roughly what I want and come out look absolutely nothing like it my hair be a horrible ashy blonde not anywhere close to the platinum blonde I request she will not do any of the pop of colour I want and even after specifically tell her I do not like blunt cut my hair have lot of straight edge she do not listen to a single thing I want and when I tell her I be unhappy with the colour she basically tell me I be wrong and I have do it this way no no I do not if I can go from Little Mermaid red to golden blonde in 1 sitting that leave my hair fine I shall be able go from golden blonde to a shade of platinum blonde in 1 sitting thanks for ruin my New Year's with 1 the bad hair job I have ever have

(a) 1 star reviews

- I really enjoy Ashley and Ami salon she do a great job be friendly and professional I usually get my hair do when I go to MI because of the quality of the highlight and the price the price be very affordable the highlight fantastic thank Ashley i highly recommend you and ill be back
- love this place it really be my favorite restaurant in Charlotte they use charcoal for their grill and you can taste it steak with chimichurri be always perfect Fried yucca cilantro rice pork sandwich and the good tres lech I have had.The desert be all incredible if you do not like it you be a mutant if you will like diabeetus try the Inca Cola
- this place be so much fun I have never go at night because it seem a little too busy for my taste but that just prove how great this restaurant be they have amazing food and the staff definitely remember us every time we be in town I love when a waitress or waiter come over and ask if you want the cab or the Pinot even when there be a rush and the staff be run around like crazy whenever I grab someone they instantly smile acknowlegde us the food be also killer I love when everyone know the special and can tell you they have try them all and what they pair well with this be a first last stop whenever we be in Charlotte and I highly recommend them
- great food and good service .... what else can you ask for everything that I have ever try here have be great
- first off I hardly remember waiter name because its rare you have an unforgettable experience the day I go I be celebrate my birthday and let me say I leave feel extra special our waiter be the best ever Carlos and the staff as well I be with a party of 4 and we order the potato salad shrimp cocktail lobster amongst other thing and boy be the food great the lobster be the good lobster I have ever eat if you eat a dessert I will recommend the cheese cake that be also the good I have ever have it be expensive but so worth every penny I will definitely be back there go again for the second time in a week and it be even good ...... this place be amazing

(b) 5 star reviews

A structured self-attentive sentence embedding

Lin, Z., Feng, M., Santos, C. N. D., Yu, M., Xiang, B., Zhou, B., Bengio, Y. (2017)

Advantages:

The final sentence embedding M have direct access to all hidden states, consequently the LSTM is allowed to focus on shorter term context. The long term dependencies are managed by the attention mechanism.

Concerns

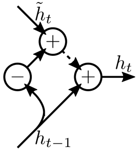
The experiments do not fully support the claim that this model will improve end task performance over more standard attention mechanisms.

Zoneout: Regularizing RNNS by randomly preserving hidden activations

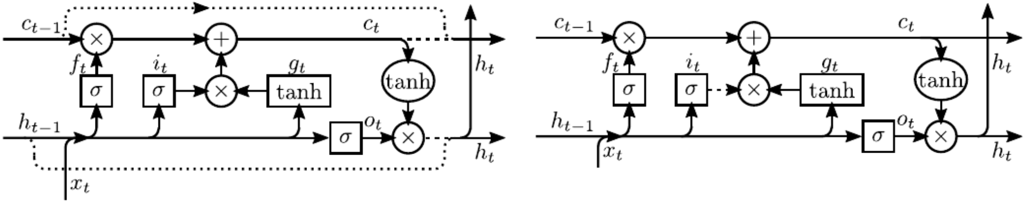
Krueger, D., Maharaj, T., Kramar, J., Pezeshki, M., Ballas, N. Courville, A. (2016)

Motivation & Approach:

zoneout stochastically forces some hidden units to maintain their previous values. Like dropout, zoneout uses random noise to train a pseudo-ensemble, improving generalization.



Comparison to recurrent dropout



Zoneout: Regularizing RNNS by randomly preserving hidden activations

Krueger, D., Maharaj, T., Kramar, J., Pezeshki, M., Ballas, N. Courville, A. (2016)

Performance improvement

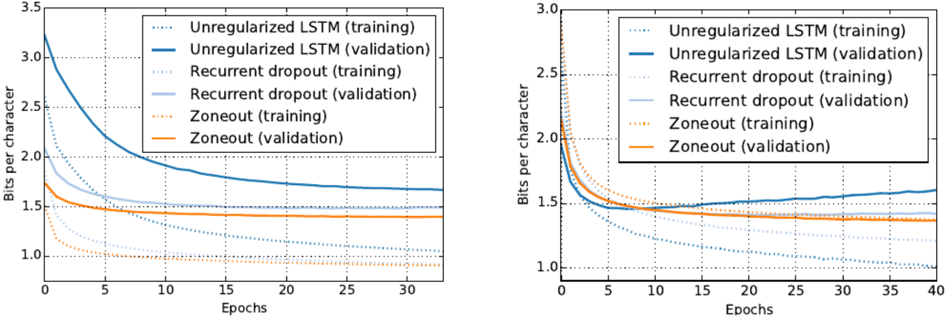


Figure 4: Training and validation bits-per-character (BPC) comparing LSTM regularization methods on character-level Penn Treebank (left) and Text8. (right)

Thank you!