

A Structured Self-attentive Sentence Embedding

Xing Wang

2018-06-25

- Motivations

- Model

- Experiments

- Motivations

- Model

- Experiments

- Sentence embedding:
 - universal sentence embeddings(unsupervise)
 - SkipThought vectors, ParagraphVectusuallyor, recursive auto-encoders, Sequential Denoising Autoencoders, FastSent
 - models trained specifically for a certain task(supervised)
 - RNN CNN
- One generally finds that specifically trained sentence embeddings perform better than generic ones, although generic ones can be used in a semi-supervised setting, exploiting large unlabeled corpora. (supervised)

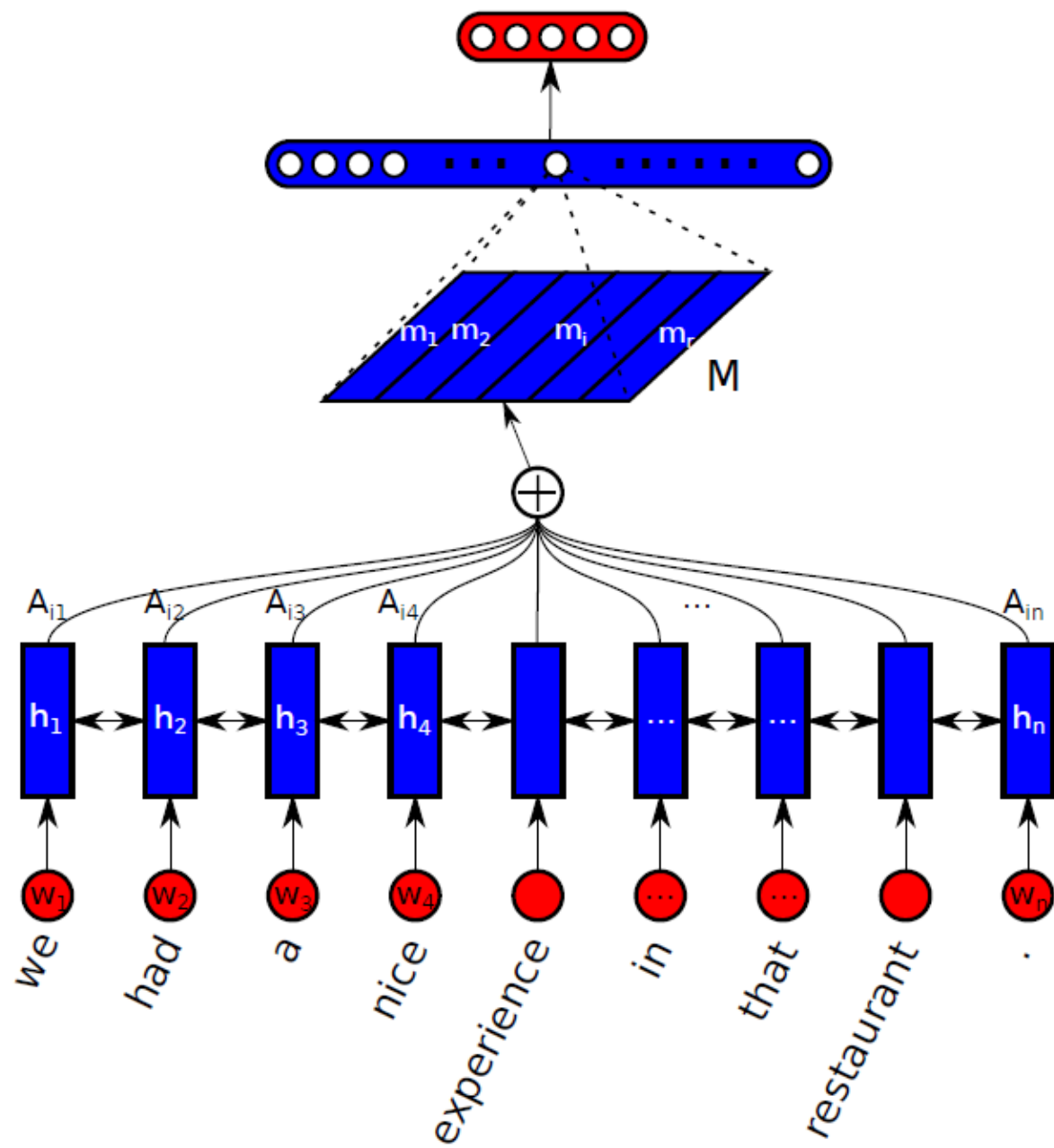
- A common approach in many of the aforementioned methods consists of creating a simple vector representation by using
 1. the final hidden state of the RNN or the max (or average) pooling from either RNNs hidden states
 2. convolved n-grams.
- We hypothesize that carrying the semantics along all time steps of a recurrent model is relatively hard and not necessary.

- Motivations

- Model

- Experiments

- We propose a self-attention mechanism for these sequential models to replace the max pooling or averaging step. Different from previous approaches, the proposed self-attention mechanism allows **extracting different aspects of the sentence** into multiple vector representations.
- self-attention → multiple feature extractors



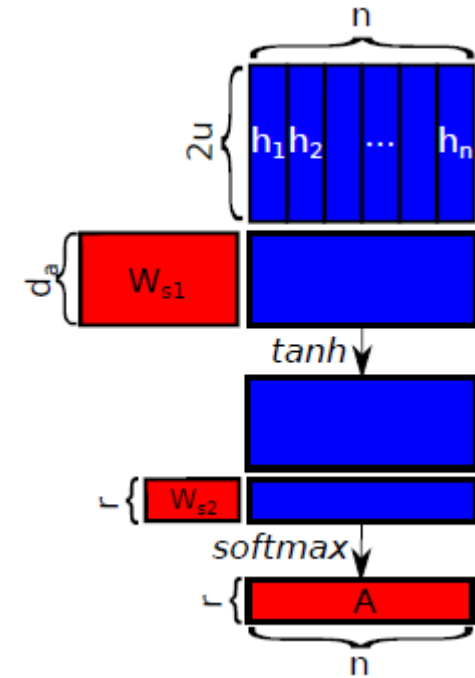
- Single feature extractor:

$$\mathbf{a} = \text{softmax}(\mathbf{w}_{s2} \tanh(W_{s1} H^T))$$

- Multiple feature extractors:

$$A = \text{softmax}(W_{s2} \tanh(W_{s1} H^T))$$

$$M = AH$$



- Penalization term:
 - to encourage the diversity of summation weight vectors across different hops of attention

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

- Here $\|_F$ stands for the Frobenius norm of a matrix.

it ' s an interesting phenomena . Not sure what the spammers get from
it . If you comment on Fastco you will get a lot of mail-replies spam .

it ' s an interesting phenomena . Not sure what the spammers get from
it . If you comment on Fastco you will get a lot of mail-replies spam .

it ' s an interesting phenomena . Not sure what the spammers get from
it . If you comment on Fastco you will get a lot of mail-replies spam .

it ' s an interesting phenomena . Not sure what the spammers get from
it . If you comment on Fastco you will get a lot of mail-replies spam .

it ' s an interesting phenomena . Not sure what the spammers get from
it . If you comment on Fastco you will get a lot of mail-replies spam .

it ' s an interesting phenomena . Not sure what the spammers get from
it . If you comment on Fastco you will get a lot of mail-replies spam .

(a)

it ' s an interesting phenomena . Not sure what the spammers get from
it . If you comment on Fastco you will get a lot of mail-replies spam .

(c) without penalization

it ' s an interesting phenomena . Not sure what the spammers get from
it . If you comment on Fastco you will get a lot of mail-replies spam .

it ' s an interesting phenomena . Not sure what the spammers get from
it . If you comment on Fastco you will get a lot of mail-replies spam .

it ' s an interesting phenomena . Not sure what the spammers get from
it . If you comment on Fastco you will get a lot of mail-replies spam .

it ' s an interesting phenomena . Not sure what the spammers get from
it . If you comment on Fastco you will get a lot of mail-replies spam .

it ' s an interesting phenomena . Not sure what the spammers get from
it . If you comment on Fastco you will get a lot of mail-replies spam .

it ' s an interesting phenomena . Not sure what the spammers get from
it . If you comment on Fastco you will get a lot of mail-replies spam .

(b)

it ' s an interesting phenomena . Not sure what the spammers get from
it . If you comment on Fastco you will get a lot of mail-replies spam .

(d) with 1.0 penalization

Figure 3: Heat maps for 2 models trained on Age dataset. The left column is trained without the penalization term, and the right column is trained with 1.0 penalization. (a) and (b) shows detailed attentions taken by 6 out of 30 rows of the matrix embedding, while (c) and (d) shows the overall attention by summing up all 30 attention weight vectors.

- Motivations
- Model
- Experiments

- Tasks:

1. Author profiling: Author Profiling dataset(Age dataset)
2. Sentiment analysis: Yelp dataset
3. Textual entailment: SNLI corpus

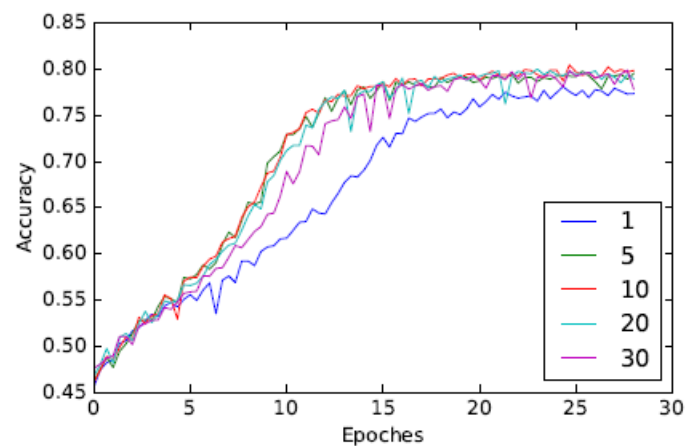
Table 1: Performance Comparison of Different Models on Yelp and Age Dataset

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%

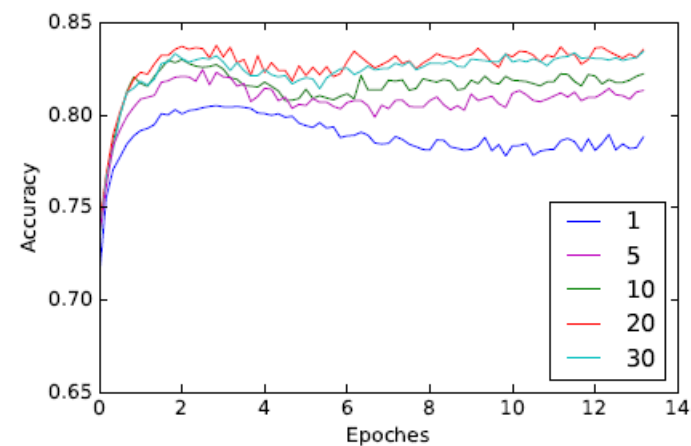
Table 2: Test Set Performance Compared to other Sentence Encoding Based Methods in SNLI Dataset

Model	Test Accuracy
300D LSTM encoders (Bowman et al., 2016)	80.6%
600D (300+300) BiLSTM encoders (Liu et al., 2016b)	83.3%
300D Tree-based CNN encoders (Mou et al., 2015a)	82.1%
300D SPINN-PI encoders (Bowman et al., 2016)	83.2%
300D NTI-SLSTM-LSTM encoders (Munkhdalai & Yu, 2016a)	83.4%
1024D GRU encoders with SkipThoughts pre-training (Vendrov et al., 2015)	81.4%
300D NSE encoders (Munkhdalai & Yu, 2016b)	84.6%
Our method	84.4%

- EFFECT OF MULTIPLE VECTORS



(a)



(b)

Figure 5: Effect of the number of rows (r) in matrix sentence embedding. The vertical axes indicates test set accuracy and the horizontal axes indicates training epochs. Numbers in the legends stand for the corresponding values of r . (a) is conducted in Age dataset and (b) is conducted in SNLI dataset.

• Visualization

- if I can give this restaurant a 0 I will 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

- 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 acknowledge 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

Figure 2: Heatmap of Yelp reviews with the two extreme score.

