DisSent: Learning Sentence Representations from Explicit Discourse Relations

Allen Nie --------------------- Department of Computer Science
Erin D. Bennett ------------- Department of Computer Science
Noah D. Goodman --------- Department of Psychology
Stanford University

Presented by: Pere-Lluís Huguet Cabot

Motivation

We have models like InferSent which worked well for sentence embeddings and are simple and straightforward.

But they need a lot of annotated data.

The are also models that don't need annotated data.

But they need a lot of data and can be complex and slow to train.

Using a self-supervised approach can take the best from both worlds. How?

Discourse markers

Keys

- An explicit discourse marker dataset generator.
- Use an unsupervised method to get sentence embeddings.
- Be good at it.
- Provide a new task and dataset.

What is an Explicit Discourse Relation?

[I wore a jacket] because [it was cold outside].

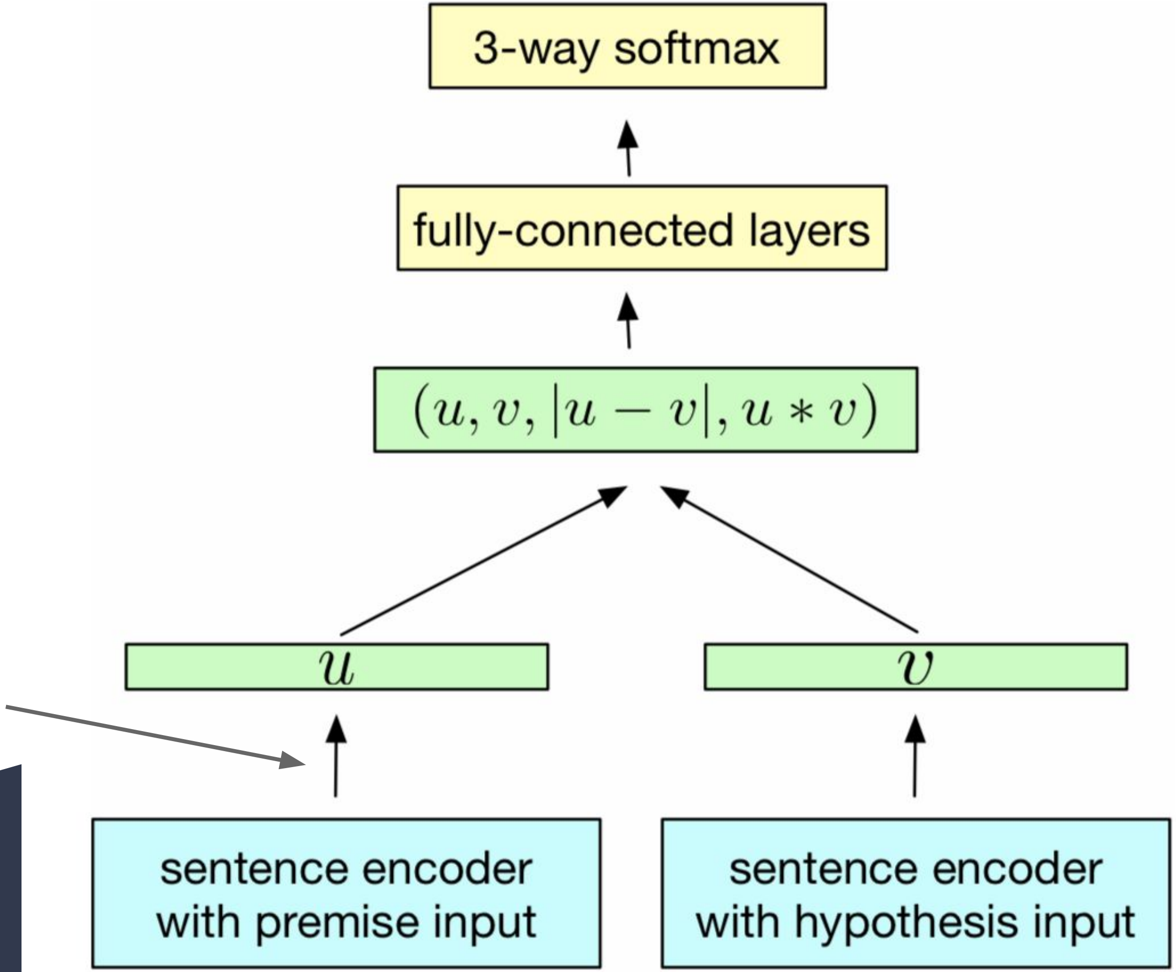
S1 *marker* S2

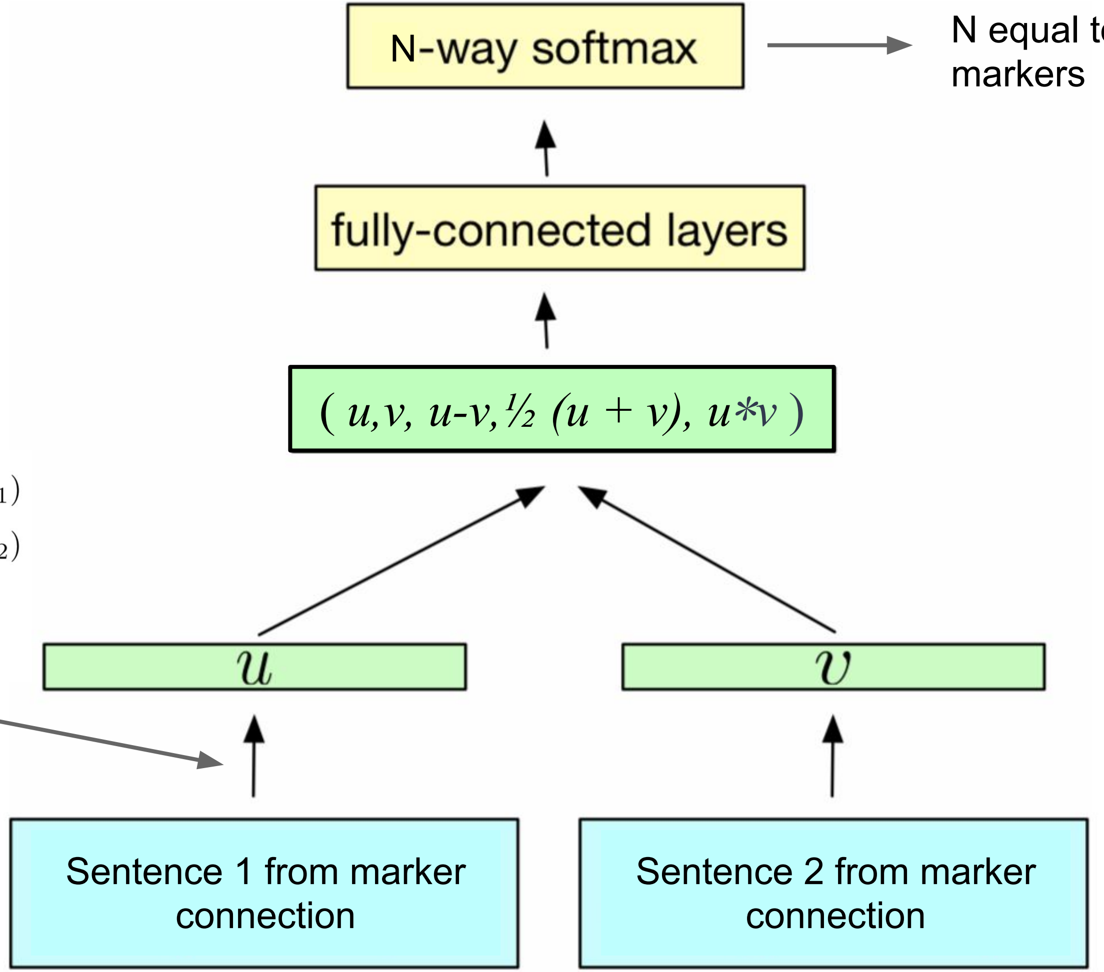
Because [it was cold outside], [I wore a jacket].

marker S2 S1

Bidirectional LSTM:

$$\overrightarrow{h_t} = \text{LSTM}_t(w_1, ..., w_T | \theta_1)$$
$$\overleftarrow{h_t} = \text{LSTM}_t(w_T, ..., w_1 | \theta_2)$$
$$h_t = [\overrightarrow{h_t}; \overleftarrow{h_t}]$$
$$s_i = \text{MaxPool}(h_1, ..., h_T)$$


3-way softmax
fully-connected layers
$(u, v, |u - v|, u * v)$
u
v
sentence encoder with premise input
sentence encoder with hypothesis input



N equal to number of markers

$$\overrightarrow{h_t} = \text{LSTM}_t(w_1, ..., w_T|\theta_1)$$
$$\overleftarrow{h_t} = \text{LSTM}_t(w_T, ..., w_1|\theta_2)$$
$$h_t = [\overrightarrow{h_t}; \overleftarrow{h_t}]$$
$$s_i = \text{MaxPool}(h_1, ..., h_T)$$

Dataset (BookCorpus)

Same Sentence (**SS**):

- [I wore a jacket] <u>because</u> [it was cold outside].

Immediate Predecessor (**IPS**):

- <u>Because</u> [it was cold outside], [I wore a jacket].

Non-adjacent previous sentence (**NAPS**):

- [I wore a jacket], which my sister gave to me, <u>because</u> [it was cold outside].

Sentence that follows (**FS**):

- [It was cold outside]. <u>So</u> [I wore a jacket]

1. Choice of Discourse Markers
   a. At least 1% in the overall corpus
2. Stanford CoreNLP dependency parser
   a. 91% of instances
3. Length-based Filtering
   a. Less than 5
   b. More than 50
   c. Ratio of more than 5

Outcome:

4,706,292 pairs of sentences for 15 discourse markers.

Train: 90%
Validation: 5%
Test: 5%

With no supervision!

=)

Marker	Extracted Pairs	Percent (%)
and	818,634	21.1
as	761,330	19.6
when	552,540	14.2
but	508,648	13.1
if	491,394	12.6
before	268,787	6.9
while	120,231	3.1
because	116,444	3.0
after	84,330	2.2
though	61,023	1.6
so	57,816	1.5
although	13,933	0.4
still	11,125	0.3
also	10,026	0.3
then	8,414	0.2
Total	4,706,292	100.0

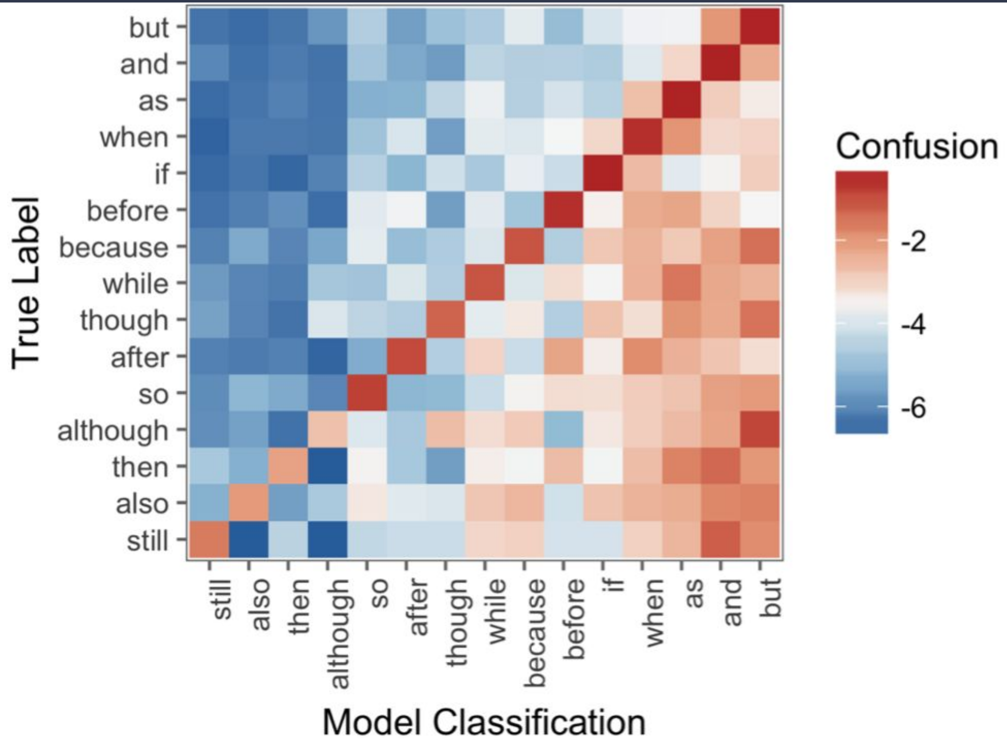
Table 1: Number of pairs of sentences extracted from BookCorpus for each discourse marker and percent of each marker in the resulting dataset.

Experiments

- Implicit vs. Explicit Prediction Task
- Implicit Relation Prediction Task
- *SentEval* (Conneau et al. (2017)):
  - sentiment analysis (MR, SST)
  - question-type (TREC)
  - product reviews (CR)
  - subjectivity-objectivity (SUBJ)
  - opinion polarity (MPQA)
  - entailment (SICK-E)
  - relatedness (SICK-R)
  - Paraphrase detection (MRPC)
- New Task: DIS, with a new dataset.

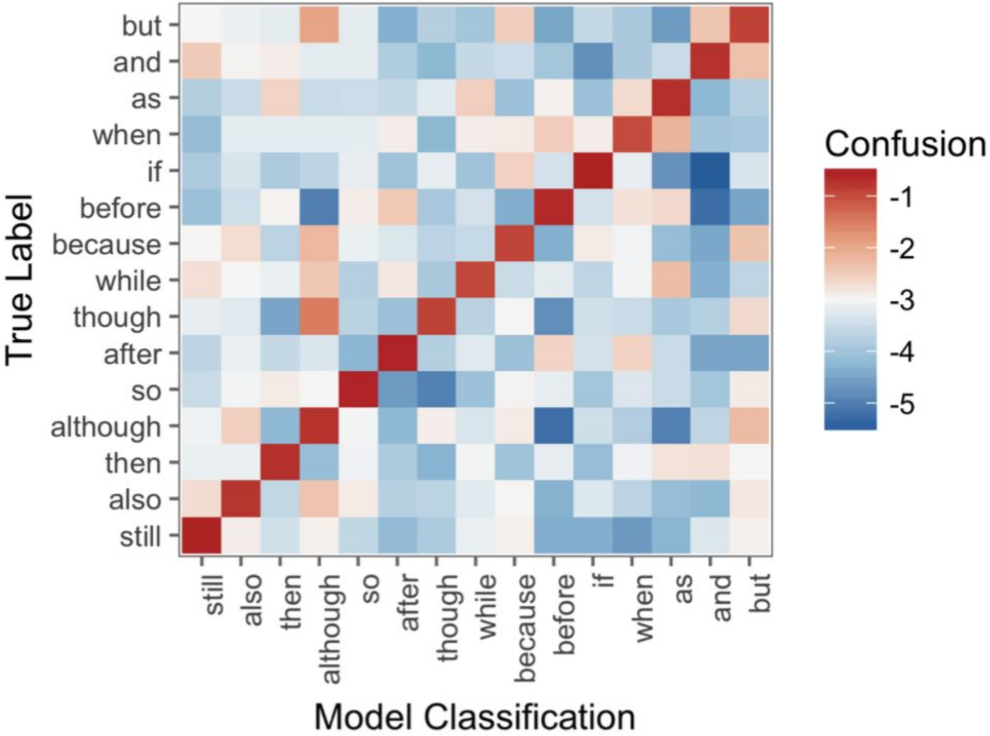
SDG, learning rate 0.1 with factor 5 annealing, 20 epochs, no dropout. Max pooling. 4096 hidden state size.

Results



but
and
as
when
if
before
because
while
though
after
so
although
then
also
still
True Label
still
also
then
although
so
after
though
while
because
before
if
when
as
and
but
Model Classification
Confusion
-2
-4
-6

Results



True Label
but
and
as
when
if
before
because
while
though
after
so
although
then
also
still
Model Classification
still
also
then
although
so
after
though
while
because
before
if
when
as
and
but
Confusion
-1
-2
-3
-4
-5

Marker	All	Books 8	Books 5
and	0.78 / 0.72	0.78 / 0.78	0.79 / 0.81
but	0.73 / 0.71	0.79 / 0.72	0.80 / 0.75
because	0.36 / 0.45	0.37 / 0.50	0.38 / 0.55
if	0.75 / 0.79	0.80 / 0.78	0.81 / 0.81
when	0.62 / 0.61	0.74 / 0.71	0.77 / 0.77
so	0.48 / 0.49	0.46 / 0.56	—
though	0.30 / 0.48	0.39 / 0.61	—
before	0.61 / 0.65	0.64 / 0.77	—
as	0.77 / 0.68	—	—
while	0.36 / 0.46	—	—
after	0.42 / 0.55	—	—
although	0.07 / 0.24	—	—
still	0.21 / 0.42	—	—
also	0.14 / 0.36	—	—
then	0.12 / 0.31	—	—
Overall	67.5	73.5	77.3

Table 4: **Training task performance:** Test recall / precision for each discourse marker on the classification task, and we report overall accuracy.

Model	IMP	IVE
DisSent Books 5†	40.7	86.5
DisSent Books 8†	41.4	**87.9**
DisSent Books ALL†	**42.9**	87.6
InferSent (Conneau et al., 2017)	38.4	84.5
Patterson and Kehler (2013)	—	86.6
Word Vectors (Qin et al., 2017)	36.9	74.8
Lin et al. (2009) + Brown Cluster	40.7	—
Adversarial Net (Qin et al., 2017)	**46.2**	—

Table 5: **Discourse Generalization Tasks using PDTB:** Following the metric used in these literature, we report overall test accuracy for sentence embedding models, as well as baselines and state of the art for these task.

IMP = Implicit Prediction

IVE = Implicit VS Explicit

Model	MR	CR	SUBJ	MPQA	SST	TREC	SICK-R	SICK-E	MRPC	DIS
Self-supervised training methods										
DisSent Books 5†	80.2	85.4	93.2	90.2	82.8	91.2	0.845	83.5	76.1	75.7
DisSent Books 8†	79.8	85.0	93.4	90.5	83.9	93.0	0.854	83.8	76.1	80.2
DisSent Books ALL†	80.1	84.9	93.6	90.1	84.1	93.6	0.849	83.7	75.0	79.9
Disc BiGRU	—	—	88.6	—	—	81.0	—	—	71.6	—
Unsupervised training methods										
FastSent	70.8	78.4	88.7	80.6	—	76.8	—	—	72.2	
FastSent + AE	71.8	76.7	88.8	81.5	—	80.4	—	—	71.2	—
Skipthought	76.5	80.1	93.6	87.1	82.0	92.2	0.858	82.3	73.0	70.1
Skipthought-LN	79.4	83.1	93.7	89.3	82.9	88.4	0.858	79.5	—	—
Supervised training methods										
DictRep (bow)	76.7	78.7	90.7	87.2	—	81.0	—	—	—	—
InferSent	81.1	86.3	92.4	90.2	84.6	88.2	0.884	86.1	76.2	65.4
Multi-task training methods										
LSMTL	82.5	87.7	94.0	90.9	83.2	93.0	0.888	87.8	78.6	—

SentEval Tasks + DIS

Conclusion and Personal thoughts

Self-supervised methods provide a huge advantage in terms of getting a dataset.

Provides a great tool to use in other tasks.

Use in other languages.

Limitations of evaluation.

Provides a new task and dataset to evaluate.

This is how we got models for word embeddings, which means this can be a good path for sentence embeddings.

Future

We could check for more insight on the embeddings by checking which neurons activates each marker.

Use other relations for the same purpose. Sentence order has been used. Maybe use other relations like Punctuation marks. Although it may not comprise semantic meaning it may show other interesting structure relations.

GAME OF THRONES SPOILER

-0.013567229,-0.05065919,0.026242753,0.052643698,0.09929497,0.08925408,-0.08122494,0.003478388,-0.85253406,-0.2620143,0.008806467,0.09804541,0.15074195,0.11676946,0.14856668,-0.08855793,-0.08078795,-0.14169352,0.61937505,0.16016422,-0.4697486,0.0046804794,-0.081460096,-0.32779753,-0.2291475
1,0.090517186,-0.36738896,0.18631452,-0.5247035,1.0362767,-0.30697438,-0.10484952,-0.21930106,-0.28162715,-0.09349052,-0.21129334,-0.0120821195,0.12975086,-0.02454465,0.45860133,-0.24405783,0.05987305,0.2457904,-0.15731245,-0.021265069,-0.23824544,-0.102207065,-0.23869407,-0.14823188,0.2292996
2,0.23941462,-0.1298147,-0.3582244,0.33866784,0.40356383,-0.30415282,-0.20082545,-0.061752763,-0.50107175,-0.3202425,0.16905165,0.17953634,0.21655901,-0.019527068,0.6508384,0.56031513,0.70327216,0.34607401,-0.5072662,-0.12874098,-0.0542742,-0.55966586,-0.19196606,-0.30829886,0.23610894,0.75482
893,0.2855167,-0.025399879,0.027368337,0.14721592,-0.15505421,0.42171666,0.17714985,0.44994065,0.5194573,0.112058796,-0.059899908,-0.4409076,-0.381588,-0.19035298,0.25561062,-0.8877832,0.29758528,0.5372079,0.44139048,0.4177531,-0.1384448,0.41534495,-0.2312433,-0.17496711,0.5355561,0.06421471
4,-0.28472367,-0.050444465,0.3086579,-0.0987074,-0.21141553,0.7123861,-0.078357115,-0.70241505,0.40909755,0.33751473,1.1192895,-0.3489803,0.38039863,-0.320154,0.7176334,0.08759117,0.38723007,-0.41488993,0.6453178,-0.20598467,0.4546781,0.30398184,0.08010384,0.20020127,-0.31642053,0.21925215,-0.
21859728,0.6547562,-0.28924307,0.2572041,-0.40313137,-0.2441295,-0.07712171,0.36639357,-0.16890967,0.24199246,0.40213227,0.10417066,0.1873161,-0.041245062,-0.5248487,0.41789317,0.5107836,-0.2808315,-0.18040572,0.5712464,-0.090254374,0.3098191,0.15854931,0.26031843,-0.071412824,0.3081934,0.144
37549,-0.23776038,0.85451335,0.14323904,0.23283549,-0.11750856,-0.0026584864,0.5396247,0.3182182,-0.31495655,-0.329856,0.05323583,0.01372839,-0.58371717,-0.39536926,-0.15189956,0.033743013,0.10047538,0.08179001,-0.1608686,0.3954918,0.22391474,-0.4562075,-0.04681717,-0.43991813,0.17534949,0.68
13515,-1.7237368,-0.17808644,-0.014176766,0.2439649,-0.6577654,-0.39612046,0.016654292,0.14051627,-0.46269822,0.3093409,-0.81227493,-0.7522037,0.20235674,0.16458948,-0.4550797,0.026168803,-0.46335268,-0.7888599,0.6207516,0.19720602,0.24633189,0.43491688,-0.4187154,-0.5752451,0.12635942,0.16038
725,0.5781595,0.15301277,0.2987931,1.9721454,0.0060143187,-0.21664761,-0.47025165,0.10977492,0.040216077,-0.3159784,0.14706405,-0.08887235,0.20192312,-0.23871444,-0.21595979,-0.2945604,0.117745586,-0.51000816,0.56093115,-0.2718948,0.21428911,-0.28207597,0.088396885,-0.2537338,-0.5558467,0.479
06923,0.23446782,-0.59654415,0.1574558,0.2176613,0.09388864,0.07503256,0.55119723,0.1337056,0.20113982,0.9707293,-0.66720515,-0.70103484,-0.58755994,-0.12414279,-0.27769002,0.012994389,-0.47620758,0.58015484,-0.84388113,-0.53960186,-0.33638227,0.39691684,-0.7515614,0.26107204,-0.4731638,0.408
5149,-0.062450934,0.727434,-0.73047304,0.2574426,-0.22234726,-2.380709,-0.35858902,0.20100176,-0.39345813,-0.27946115,0.24768306,0.34660706,0.09827101,0.27791145,-0.36231744,0.30564722,0.31102365,-0.19886975,-0.3313748,0.11875386,-0.19982117,0.30632263,0.7089544,0.08751067,-0.17347968,0.62839
01,-0.29285374,0.57136,0.1264425,-0.057080727,0.1353751,0.18814199,0.033830922,0.33306798,0.5716856,0.24031867,0.28856638,0.42902613,-0.39699087,0.16204868,-0.19545718,-0.05493782,-0.15302525,-0.14120354,-0.2891007,-0.6485002,0.27439573,-0.12587066,0.3345182,-0.022358514,-0.067395456,0.216371
3,-0.15242364,-0.24634737,0.6946778,-0.21968204,0.07775829,0.15796791,0.022294438,0.36533502,-0.026487017,-0.025505254,-0.58581704,0.077938884,-0.6220388,-0.29352462,-0.14735772,-0.030880997,-0.20718126,-0.51679873,-0.09976696,-0.13162075,0.13954483,0.356338,-0.33060625,-0.201564,-0.31044,0.5433
521,-0.014899899,0.31027484,-0.608283,-0.058918685,-0.12277957,-0.09632641,-0.044303704,-0.053677622,0.12087076,0.017280081,-0.07284997,-0.4408004,-0.3659861,-0.0203361,-0.24863933,-0.37645376,-0.03625061,0.110526204,0.076905884,0.3673236,0.43492153,0.030704001,0.27334726,-0.067678794,-0.16382
714,-0.050075848,-0.09348428,0.18567288,0.027650833,-0.077441655,-0.38228664,-0.36044756,-0.47441182,0.04635298,0.6299816,0.048936486,0.27953243,-0.3192648,0.64177203,-0.43803796,0.20563759,0.101640105,-0.13403776,-0.43810606,-0.5424482,0.2270542,-0.04805426,-0.17323552,-0.4797221,-0.4362146,-0
.40282103,-0.22092648,-0.074442886,0.17968233,0.31355026,-0.22383815,0.07900626,0.37349617,-0.26304093,-0.12079946,-0.059516076,0.20732848,-0.15283495,-0.18321224,-0.2954508,0.25047863,0.11104212,0.2644535,0.2751526,-0.37018728,0.33807978,-0.001990487,-0.16665815,0.005252103,0.43934083,0.339
75777,-0.5638128,0.37753022,-0.045259554,-0.09722654,-0.0011407137,-0.[illegible]5265227,[illegible]07276921,-0.17530107,0.079529665,0.46332827,0.[illegible]4107329,0.19389947,-0.47803107,0.13143796,-0.9214017,0.13830298[illegible]0.15862763,-0.22200112,0.18135953,0.4717966,-0.11934585,0.24977225,-0.04557711,0.04823308,
0.8517743,-0.2639874,-0.17508394,0.19251184,0.09203496,-0.07520742,0.1[illegible]576,0.09929004,-0.133[illegible]706,0.41288865,0.6949532,-0.25324807,0.27998304,0.6319554,0.073480554,-0.014425485,-0.62454486,0.24797551,-0.0899482,-0.43449637,0.53364724,0.12811327,-0.075143814,0.075170465,-0.5156
93,-0.41197118,0.30676988,0.0836344,-0.00969708,0.4082477,-0.12634821,[illegible]0.23949663,0.25845382,-0.5753502,-0.2172414,-0.50834066,0.1919155
7,0.49517986,0.37853637,-0.7254806,0.3077773,0.49529734,0.048567194,-0.[illegible]0.022360131[illegible]944,-0.4356996[illegible]0.0927113[illegible]06,0.04007976,0.42028606,0.030420056,-0.07452915,0.15102607,-0.22
08883,0.08018591,0.12293265,0.020260513,0.04069825,-0.436445,-0.1721867[illegible]73796,0.24[illegible]74906,0.247[illegible]875,0.395[illegible]793,0.17[illegible]70199,0.10[illegible]1885,0.17459913,0.018144926,-0.28878504,-0.089011006,0.008794446
5,0.68330675,-0.30618942,0.27626204,-0.22611411,-0.13915287,0.39321765,0.[illegible]79,0.067322575,-0.016783318,-0.12900059,-0.10729444,0.19350809,0.
37897244,0.037342545,0.44021764,0.36793754,-0.056617826,0.1875689,-0.13594878,-0.044731498,-0.009044786,-0.39914373,0.32652715,0.057275746,0.43344012,0.21756636,0.01395909,-0.099797495,-0.48294958,-0.014599343,0.31516796,-0.1[illegible]532279,-0.14801054,0.026674425,0.57957923,0.23638214,0.012079258
,0.35705972,0.16037911,-0.03891881,-0.05262732,-0.067384265,0.2660826,0.10901481,-0.10645475,-0.1121524,0.09177951,0.21960743,0.23096879,-0.16740698,0.12236344,-0.17640693,0.19389738,0.16584523,0.20096838,0.047981054,-0.22[illegible]934,0.099181205,0.07482973,-0.02666025,-0.2699064,0.08261917,0.2735
4714,0.25054875,-0.04777385,0.45809558,0.16712557,0.11731102,0.0949258,0.0039088526,0.6863708,-0.056384042,-0.2502056,-0.5922386,0.29500556,0.25359288,0.2491434,0.005774816,0.013555627,0.04911719,0.08147656,0.1772875,0.28199032,0.19849749,-0.07276922,-0.108136155,-0.109188996,0.28783646,0.2
4856661,-0.3656592,-0.2085958,-0.0009583831,-0.4268849,0.33065656,-0.098054886,-0.05187084,-0.1688749,-0.35807195,-0.040001277,0.22208421,0.32207108,0.0861428,0.17992307,-0.17590874,-0.008620322,-0.14354266,-0.19838251,0.13965927,0.0736056,0.09632474,-0.15529291,-0.10465231,-0.036839068,0.1380
0316,-0.064661264,-0.17407435,0.25130674,0.47882822,-0.14312087,0.36564767,-0.13234793,-0.21634014,-0.7119855,-0.1263959,-0.113274015,-0.38026524,0.005835367,0.22086488,-0.25865057,0.43748248,-0.99852633,0.06161662,-0.015304749,0.1601458,-0.19760633,0.16882707,-0.01468129,0.047846686,-0.56381
05,-0.7034913,-0.18319587,-0.3678269,0.09007352,0.015363898,0.16076943,0.9523849,-0.17218107,-0.3768424,-0.047230065,-0.062186316,-0.21647234,-0.14694779,0.17445083,0.35612068,0.08536342,-0.14407313,-0.16879378,-0.21756953,-0.13563418,0.21268104,0.033393506,-0.03315781,-0.31135598,-0.026836425,
0.011363621,0.45067084,0.42750502,-0.35585,0.8014485,0.09039483,-0.18496639,-0.33371225,-0.043529768,0.6262701,0.39348292,0.33209798,-0.28328824,0.10213134,-0.22600274,-0.2360726,0.026616434,0.22766459,0.43829957,0.23504394,-0.23925269,0.02094674,-0.3251567,0.19609888,0.25306895,0.3232316,-0.
1027172,0.05758278,-0.5486476,0.33663726,-0.18002255,-0.19177496,0.020358613,0.018467724,-0.5373531,0.0037915632,0.20551592,-0.34276852,-0.07526708,-0.49883044,0.12198558,0.019597163,-0.16714126,0.12820621,0.17606466,0.15603614,-0.31840423,0.13049853,0.17746277,-0.0066163517,0.07590871,-0.1
9115369,0.14571439,0.08397806,0.48358062,-0.66882867,0.23419265,-0.12847629,-1.7159358,-0.42551455,0.05978386,-0.008263528,0.3306922,0.25006926,0.21859604,0.03253674,0.014542679,0.051910635,-0.11918068,0.24358225,-0.13970053,0.12440154,-0.28420624,-0.08248394,-0.25092575,-0.020045126,-0.1935
5297,0.27393416,0.58600676,-0.07714129,0.3806326,0.014328927,-0.32365033,0.1645677,-0.28177914,-0.025924856,0.24767692,-0.10975654,-0.05762286,0.11779767,0.6432616,-0.18415253,-0.06853163,-0.25028265,-0.15001583,0.23356454,-1.393025,0.16361563,-0.45081457,-0.1360646,0.16260977,0.19439542,-0.23
282318,-0.12062216,-0.12753977,0.48565888,-0.42221355,0.21278799,-0.07426339,0.32345974,0.1646455,0.10614407,0.02321437,-0.25329784,0.1832513,-0.3786694,0.259442,0.25015202,0.17536557,-0.3981534,0.05638194,-1.5086559,0.029349437,-0.05991594,0.05966213,-0.10967111,-0.105465196,-0.3543478,-0.027
244909,-0.33862486,-0.0870798,-0.026211748,0.56132174,-0.40395403,0.10685888,-0.11074021,-0.058647204,0.3028984,0.05046169,0.03698826,0.09503919,-0.15625022,-0.3115803,-0.10267358,0.010712882,0.049197454,0.10978991,0.13569528,0.04174803,0.22446631,0.08756221,-0.15395011,0.13739316,0.1316683
1,0.085838675,-0.34561396,0.1641518,0.089273155,-0.1213783,0.13074599,-0.14910676,0.22454767,-0.22400069,-0.55623394,0.00967748,0.13882798,0.677718,-0.1536464,-0.3566058,0.22338782,0.30716264,0.07180151,0.029732296,-0.40260312,-0.055670887,-0.37281427,0.38001004,0.011032276,0.21705776,-0.6685
6855,-0.36869487,0.06587965,-0.0026976466,0.030155987,-0.09276434,-0.2706722,-0.32622775,-0.03910302,0.30943027,0.17664786,-0.0028777968,-0.03976197,0.08511219,-0.036930837,-0.057527874,0.22591855,-0.11930182,-0.29615065,-0.032633338,0.09178356,0.2119106,0.28338075,-0.09963337,-0.29753473,0.1
2467438,0.18189389,0.13060583,0.067038424,-0.05136313,-0.22954927,-0.22405028,-0.09096571,-0.04527704,0.023153028,-0.13105522,0.16320546,-0.18892993,0.19802165,-0.019916514,-0.09083531,-0.0035231214,-0.2501814,-0.469375,-0.45530978,-0.24605505,-0.41469994,0.08228839,-0.08304912,0.14475161,0.37
519905,-0.039661314,-0.19777346,0.06621914,-0.2615272,-0.3234518,0.18862455,0.16814749,0.21958418,0.20300777,-0.3793724,0.07084962,0.5471289,-0.08723465,0.105474286,-0.05888482,-0.07683257,0.6105545,-0.12478775,-0.016726157,-0.23881696,-0.23567675,-0.12061838,0.049196452,0.23175704,0.0483949
97,0.11563947,-0.48630086,-0.42046562,0.08007475,0.33812997,0.36662814,0.18724376,-0.047747213,0.04838747,-0.0013216883,0.4606061,0.42644775,0.0025579904,0.18130036,-0.1218195,-0.07321215,-0.3045664,0.3567824,-0.22488242,-0.11936792,0.059771907,-0.23098671,0.16963392,0.050802644,-0.05637927,
0.06459634,0.08757775,-0.24029396,0.68302566,0.066294275,-0.24274497,-0.20747693,-0.029574424,0.07611173,-0.03976336,-0.11428422,-0.03917731,-0.02262006,-0.3089127,0.12642501,-0.26746762,-0.19086081,0.13118106,-0.01236852,-0.038146198,-0.015523563,-0.15950401,-0.30840883,-0.020964345,0.4919780
8,0.48832238

Winter is coming

Thank you very much.

Feel free to ask any question
(but don't make it too hard)

ELMo
Deep Contextualized Word Representations

Victor Zuanazzi - MSc. AI
Universiteit van Amsterdam
15/04/2019

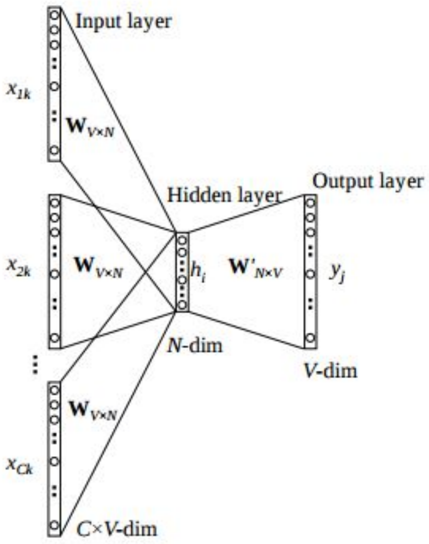
Agenda

- Word Embeddings
- (Deep) Contextualized Word Representations
- Embeddings from Language Models
- ELMo
- Experiments
- Results
- Personal Analysis
- Future Research
- Resources
- Q & A

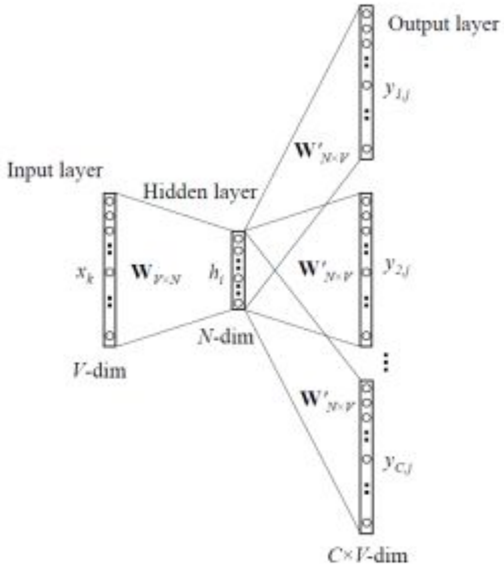
Word embeddings

- Fixed sized vectors that represent words;
- They encode some semantic and syntactic aspects of the words;
- Built using word windows or dependencies.

CBOW



Skip Gram

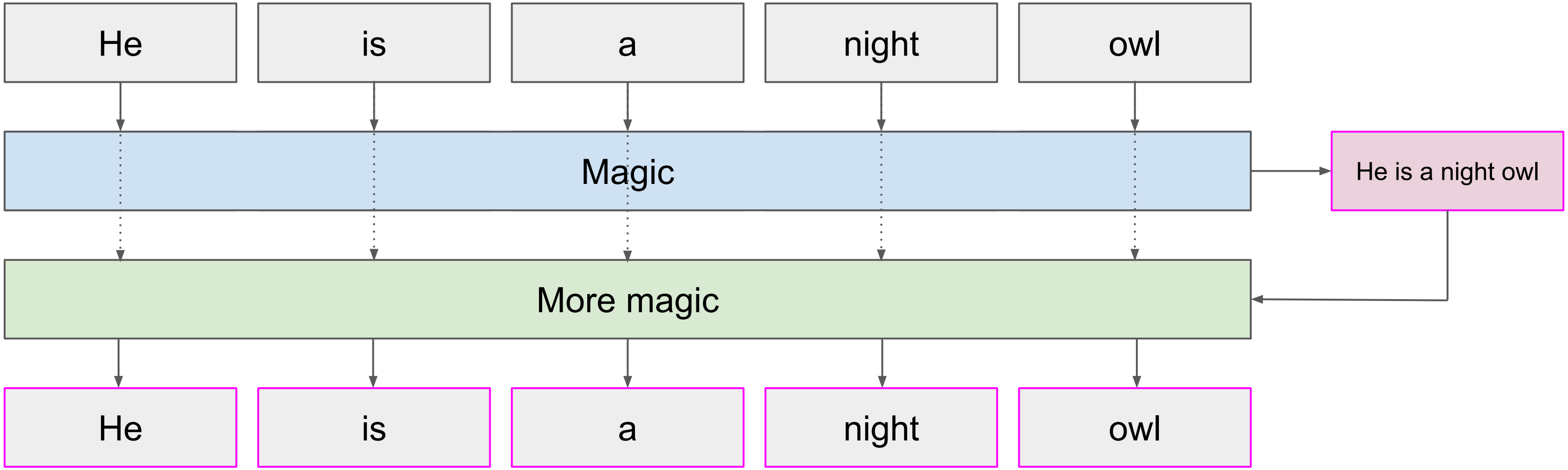


Word embeddings

- The embeddings have problems capturing polysemy;
- It is not straightforward how to use embeddings to capture sentence level semantics;
- In practice, different embeddings serve different applications.

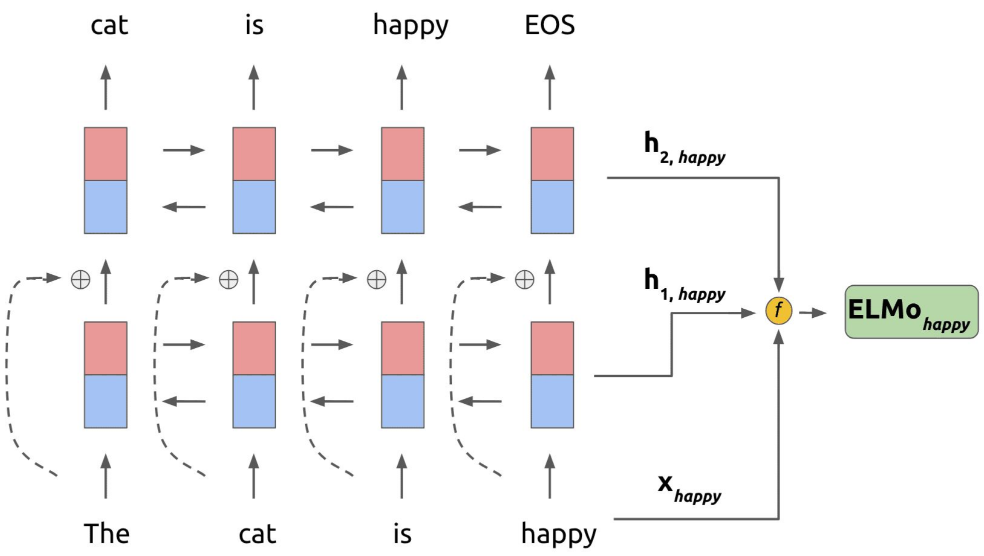
Contextualized Word Representations

- Fixed sized vectors that represent words in context;
- The same word has different representations in different contexts;
- The contextualized word representation is a function of the word embedding and the latent representation of the sentence.



Deep Contextualized Word Representations

- There is no reason for not using a deep neural net architecture and take many layers to create the (deep) contextualized word representation.



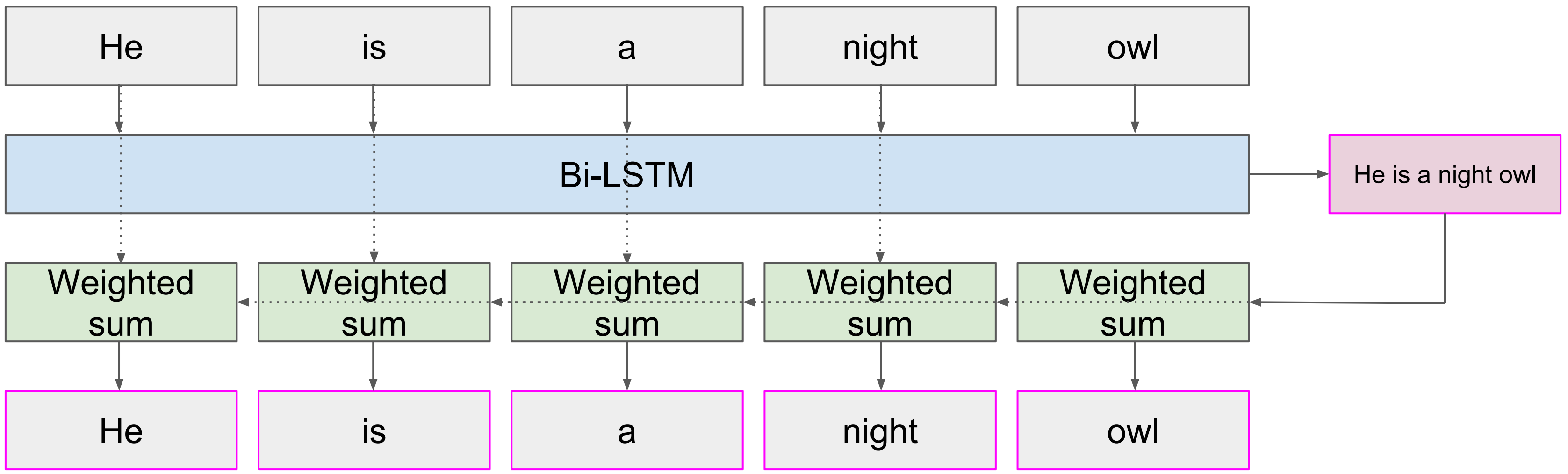
Questions

- Word Embeddings
- (Deep) Contextualized Word Representations
- Embeddings from Language Models
- ELMo
- Experiments
- Results
- Personal Analysis
- Future Research
- Resources
- Q & A

Embedding from Language Models - ELMo

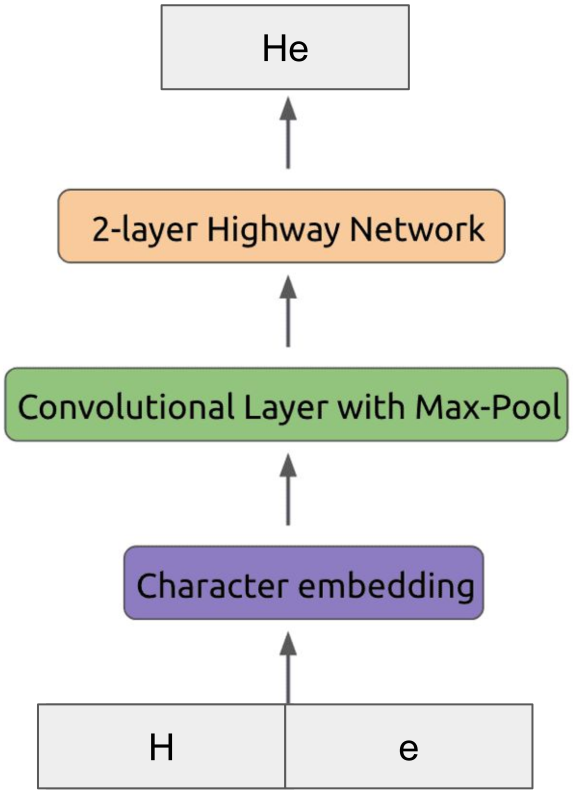
Embedding from Language Models - ELMo

- ELMo learns word representations that are dependent on context
- Learned language model: BI-LSTM



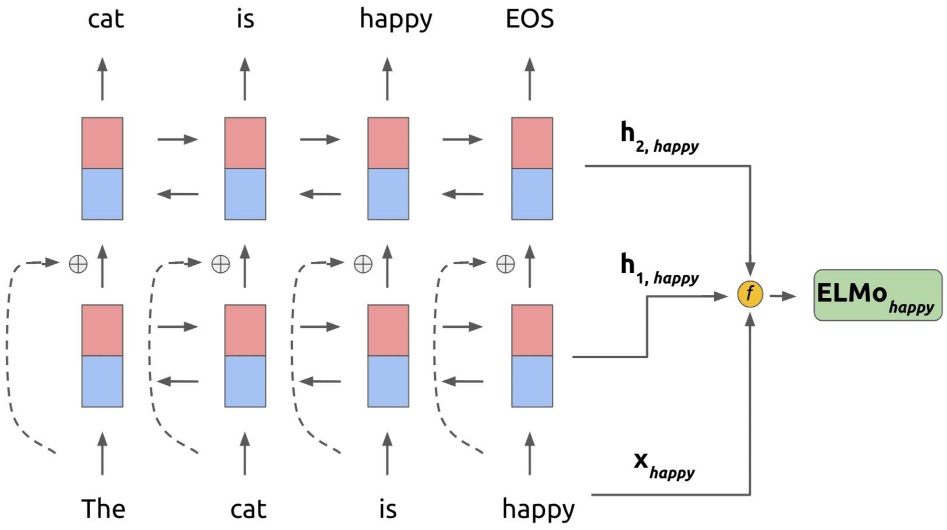
ELMo - Input

- Takes character embedding
- Convolutional Layer
- Max Pool Layer
- Highway Network



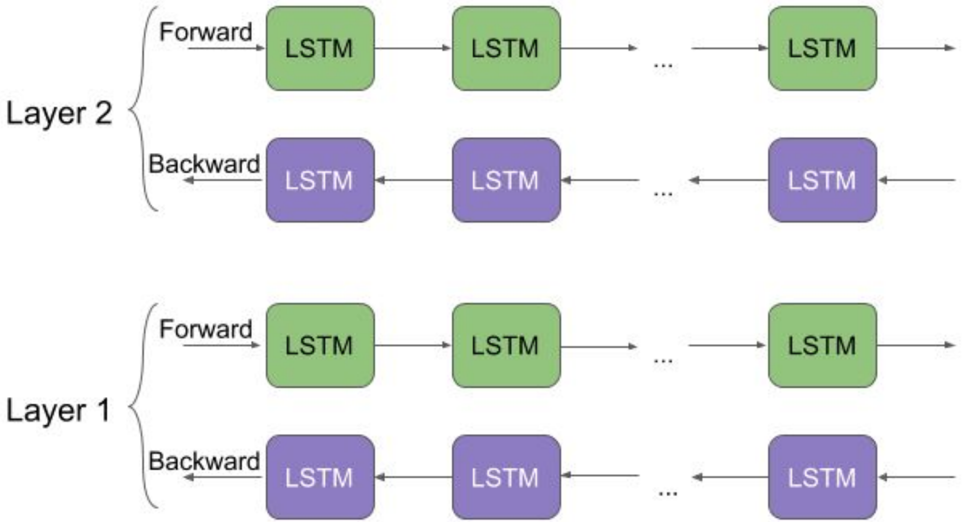
ELMo - Language Model

- The word embeddings are fed to a bi-Language Model modeled by a 2 layer bi-LSTM
- The hidden state of all layers is used together with the word embedding to create the contextualized word vector
- Trained using 30 million sentences.

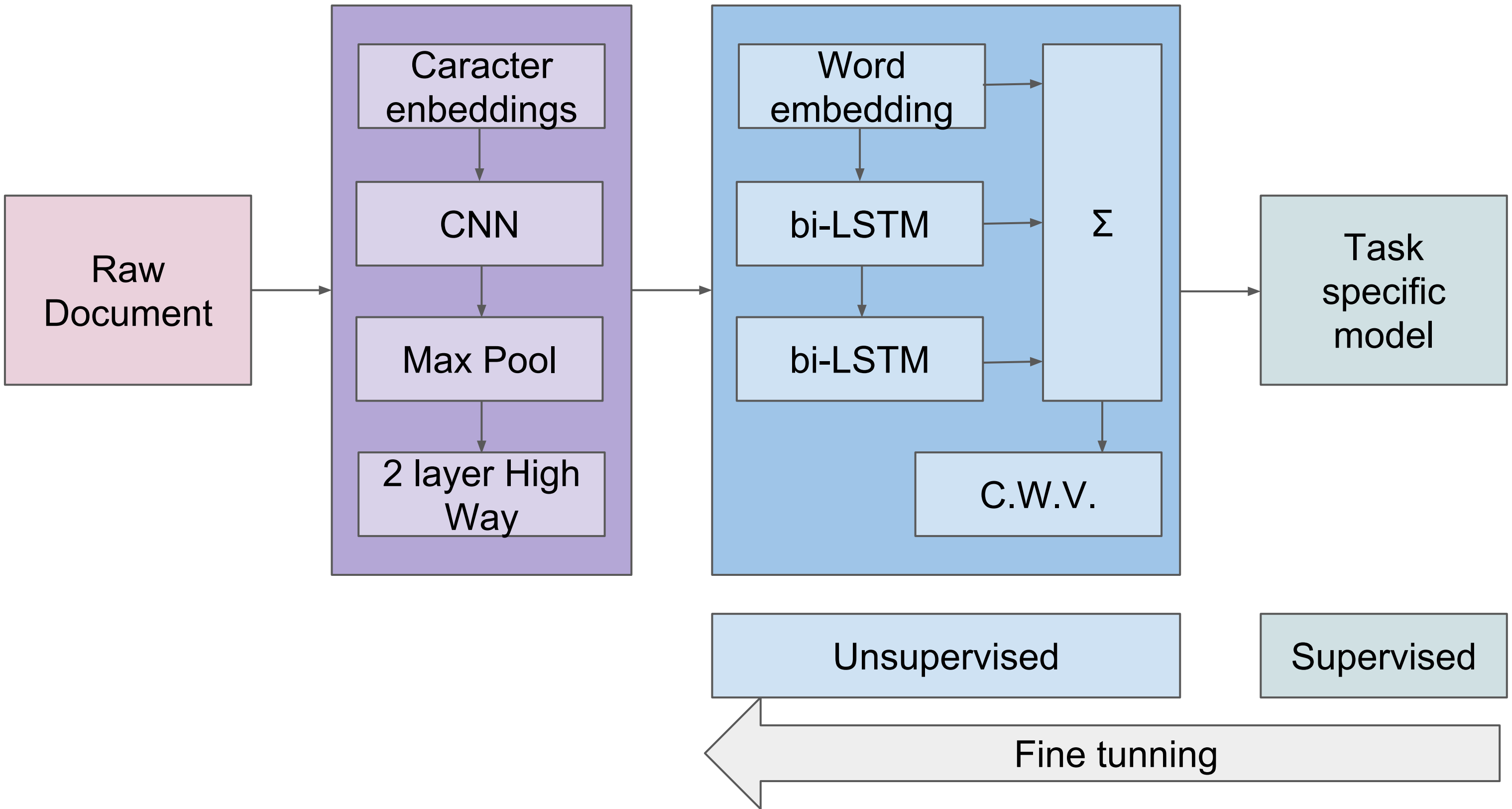


$$p(t_1, t_2, \ldots, t_N) = \sum_{k=1}^{N} \Big(\log p(t_k \mid t_1, \ldots, t_{k-1}; \Theta_x, \overrightarrow{\Theta}_{LSTM}, \Theta_s) + \log p(t_k \mid t_{k+1}, \ldots, t_N; \Theta_x, \overleftarrow{\Theta}_{LSTM}, \Theta_s) \Big) .$$

ELMo - Language Model



ELMo - supervised tasks



Questions

- Word Embeddings
- (Deep) Contextualized Word Representations
- Embeddings from Language Models
- ELMo
- Experiments
- Results
- Personal Analysis
- Future Research
- Resources
- Q & A

Experiments

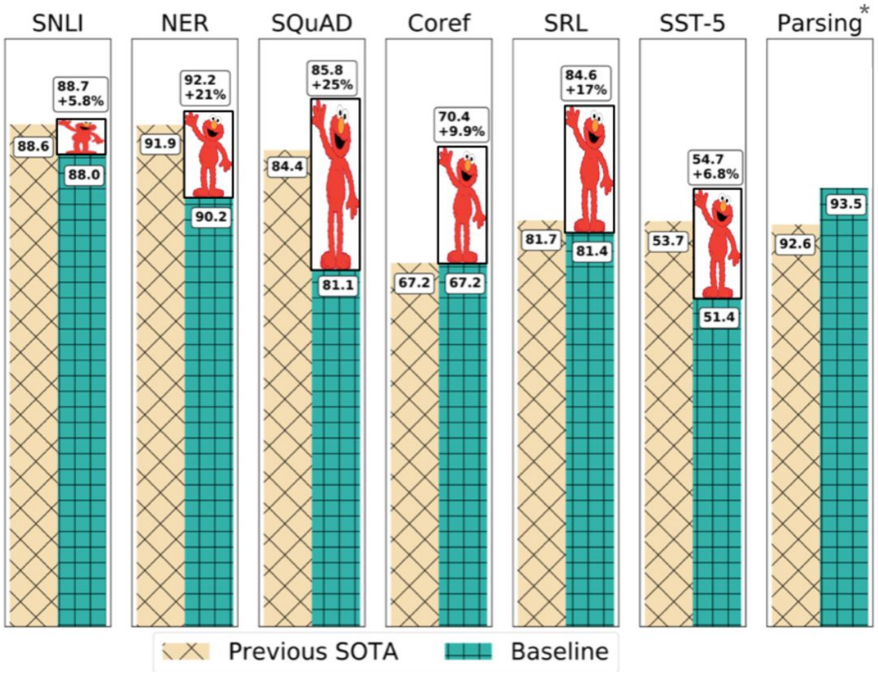
- Question Answering (SQuAD)
- Textual Entailment (SNLI)
- Semantic Role Labeling (SRL)
- Coreference Resolution (Coref)
- Named Entity Recognition (NER)
- Sentiment Analysis(SST-5)

Experiments

- Question Answering (SQuAD)
- **Textual Entailment (SNLI)**
- **Semantic Role Labeling (SRL)**
- Coreference Resolution (Coref)
- Named Entity Recognition (NER)
- Sentiment Analysis(SST-5)

Results

Task	Previous SOTA		Our baseline	ELMo + baseline	Increase (absolute/ relative)
SQuAD	Liu et al. (2017)	84.4	81.1	85.8	4.7 / 24.9%
SNLI	Chen et al. (2017)	88.6	88.0	88.7 ± 0.17	0.7 / 5.8%
SRL	He et al. (2017)	81.7	81.4	84.6	3.2 / 17.2%
Coref	Lee et al. (2017)	67.2	67.2	70.4	3.2 / 9.8%
NER	Peters et al. (2017)	91.93 ± 0.19	90.15	92.22 ± 0.10	2.06 / 21%
SST-5	McCann et al. (2017)	53.7	51.4	54.7 ± 0.5	3.3 / 6.8%



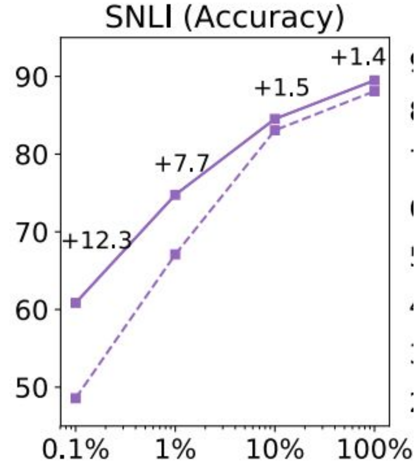

SESAME STREET


SESAME STREET

Textual Entailment - SNLI

Task	Previous SOTA		Our baseline	ELMo + baseline	Increase (absolute/ relative)
SNLI	Chen et al. (2017)	88.6	88.0	88.7 ± 0.17	0.7 / 5.8%

- Base line: ESIM;
- Small improvement;
- No significance test was reported;
- Performance is consistently better for different training set sizes.



Textual Entailment - SNLI

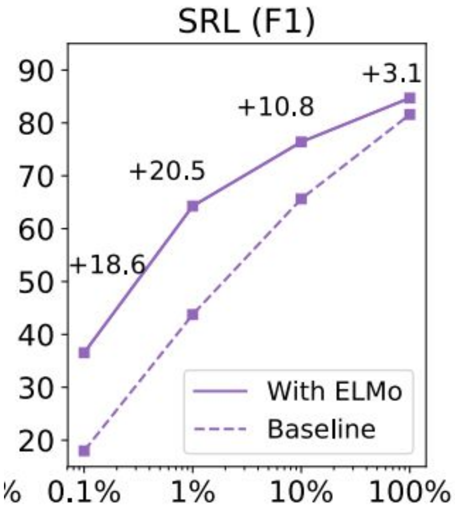
- Not current state of the art

Publication	Model	Parameters	Train (% acc)	Test (% acc)
Qian Chen et al. '16	600D ESIM + 300D Syntactic TreeLSTM (code)	7.7m	93.5	88.6
Peters et al. '18	ESIM + ELMo	8.0m	91.6	88.7
Boyuan Pan et al. '18	300D DMAN	9.2m	95.4	88.8
Zhiguo Wang et al. '17	BiMPM **Ensemble**	6.4m	93.2	88.8
Yichen Gong et al. '17	448D Densely Interactive Inference Network (DIIN, code) **Ensemble**	17m	92.3	88.9
Seonhoon Kim et al. '18	Densely-Connected Recurrent and Co-Attentive Network	6.7m	93.1	88.9
Zhuosheng Zhang et al. '18	SLRC	6.1m	89.1	89.1
Qian Chen et al. '17	KIM **Ensemble**	43m	93.6	89.1
Ghaeini et al. '18	450D DR-BiLSTM **Ensemble**	45m	94.8	89.3
Peters et al. '18	ESIM + ELMo **Ensemble**	40m	92.1	89.3
Yi Tay et al. '18	300D CAFE **Ensemble**	17.5m	92.5	89.3
Chuanqi Tan et al. '18	150D Multiway Attention Network **Ensemble**	58m	95.5	89.4
Boyuan Pan et al. '18	300D DMAN **Ensemble**	79m	96.1	89.6
Radford et al. '18	Fine-Tuned LM-Pretrained Transformer	85m	96.6	89.9
Seonhoon Kim et al. '18	Densely-Connected Recurrent and Co-Attentive Network **Ensemble**	53.3m	95.0	90.1
Xiaodong Liu et al. '19	MT-DNN	110m	96.8	**91.1**

Semantic Role Labeling - SRL

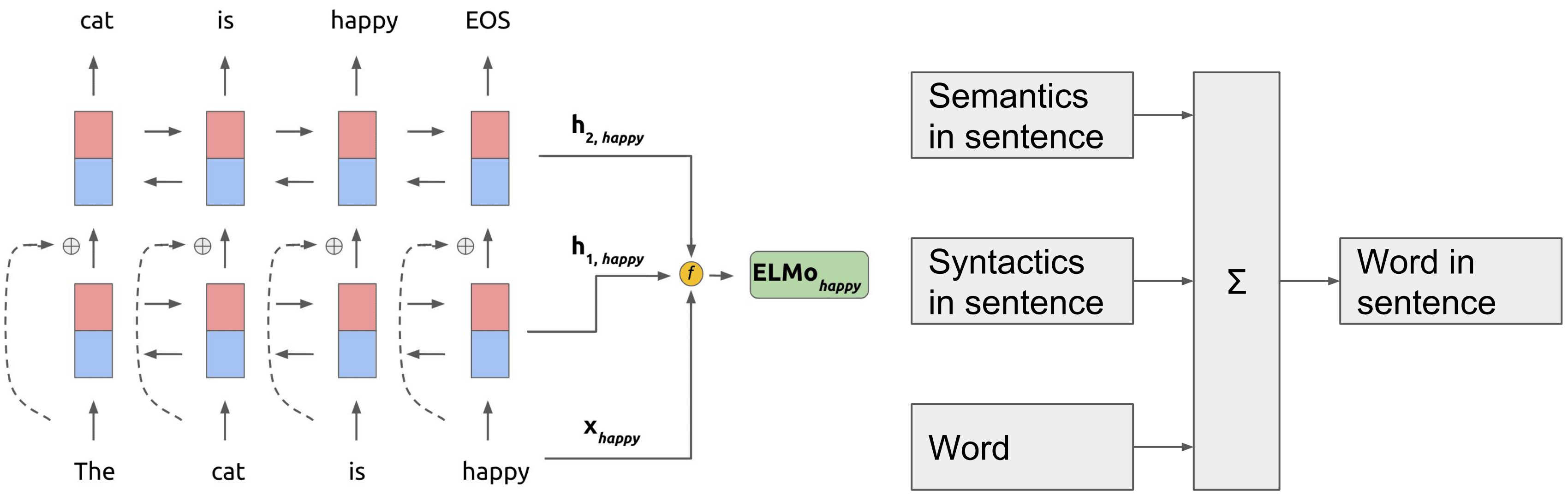
Task	Previous SOTA		Our baseline	ELMo + baseline	Increase (absolute/ relative)
SRL	He et al. (2017)	81.7	81.4	84.6	3.2 / 17.2%

- Strong improvement;
- No significance test was reported;
- Performance is consistently better for different training set sizes.



Semantic Role Labeling - SRL

- First LSTM layer captures syntactic roles;
- Second LSTM layer captures semantic roles;



Semantic Role Labeling - SRL

Model	F1	Paper / Source
He et al., (2018) + ELMO	85.5	Jointly Predicting Predicates and Arguments in Neural Semantic Role Labeling
(He et al., 2017) + ELMo (Peters et al., 2018)	84.6	Deep contextualized word representations
Tan et al. (2018)	82.7	Deep Semantic Role Labeling with Self-Attention
He et al. (2018)	82.1	Jointly Predicting Predicates and Arguments in Neural Semantic Role Labeling
He et al. (2017)	81.7	Deep Semantic Role Labeling: What Works and What's Next

Questions

- Word Embeddings
- (Deep) Contextualized Word Representations
- Embeddings from Language Models
- ELMo
- Experiments
- Results
- Future Research
- Personal Analysis
- Resources
- Q & A

Future Research

	Base model	pre-training	Downstream tasks	Downstream model	Fine-tuning
CoVe	seq2seq NMT model	supervised	feature-based	task-specific	/
ELMo	two-layer biLSTM	unsupervised	feature-based	task-specific	/
CVT	two-layer biLSTM	semi-supervised	model-based	task-specific / task-agnostic	/
ULMFiT	AWD-LSTM	unsupervised	model-based	task-agnostic	all layers; with various training tricks
GPT	Transformer decoder	unsupervised	model-based	task-agnostic	pre-trained layers + top task layer(s)
BERT	Transformer encoder	unsupervised	model-based	task-agnostic	pre-trained layers + top task layer(s)
GPT-2	Transformer decoder	unsupervised	model-based	task-agnostic	pre-trained layers + top task layer(s)

Personal Analysis

- Quite a complex architecture;
- Why bi-LSTMs? Why two layers?
- No statistical significance study made on the results;
- No need for annotated data
- “Plug and play” embeddings = Transfer Learning
- Interesting findings on syntactic and semantic modeling.

Resources

1. https://arxiv.org/abs/1802.05365
2. https://www.mihaileric.com/posts/deep-contextualized-word-representations-elmo/
3. https://www.analyticsvidhya.com/blog/2019/03/learn-to-use-elmo-to-extract-features-from-text/
4. https://towardsdatascience.com/review-highway-networks-gating-function-to-highway-image-classification-5a33833797b5
5. https://towardsdatascience.com/besides-word-embedding-why-you-need-to-know-character-embedding-6096a34a3b10
6. https://arxiv.org/pdf/1609.06038.pdf
7. https://lilianweng.github.io/lil-log/2019/01/31/generalized-language-models.html
8. http://jalammar.github.io/illustrated-bert/
9. https://nlpprogress.com/
10. http://ruder.io/nlp-imagenet/

Q & A


#FALLONTONIGHT