Bachelor Thesis

Deep Learning for Extraction of Opinion Entities

Yiyi Chen
LMU Informatik plus
Computerlinguistik

Overview

- **Sequence Labeling**: assign a label to each token in a sentence
- **Data Parsing**: deal with overlapping relations
- **Experiment**: BiLSTM(+CRF), CNN(+CRF)
- **Result**: quantitative & qualitative analysis
    - **Evaluation Metrics**: Binary/Proportional Overlap
- **Conclusion**
- **Future Work**

Sequence labeling

Labels = {O, B_H, I_H, B_T, I_T, B_O, I_O}.

Opinion entities: Opinion Expression, Opinion Holder, Opinion Target

“Our agency seriously needs equipment for detecting drugs, “ he said.

Sequence labeling

“ Our agency seriously needs equipment for detecting drugs , ” he said .

O B_H I_H B_O I_O B_T I_T I_T I_T O O B_H B_O O

Data Parsing

MPQA 2.0 corpus [Quelle: Wiebe, 2015]

Goal: To assign a label to each word, and have as many words labeled as possible.

Partial Overlapping:

The two countries have each accused the other of backing dissidents.

1. {"The two countries have each", "have each accused", "the other"}
2. {"backing", "dissidents"}

The two countries have each accused the other of backing dissidents.
B_H I_H I_H B_O I_O I_O

Data Parsing

Sub-relations:

… he was willing to consider Saudi Arabia’s request on a case-by-case basis…

1. {“he”, “was willing to”, “consider Saudi Arabia’s request on a case-by-case basis”}
2. {“he”, “consider“}
3. {“Saudi Arabia”, “request on”}

… he was willing to consider Saudi Arabia’s request on a case-by-case basis…

B_T I_T I_T I_T I_T I_T I_T I_T

Average number of words per opinion entity

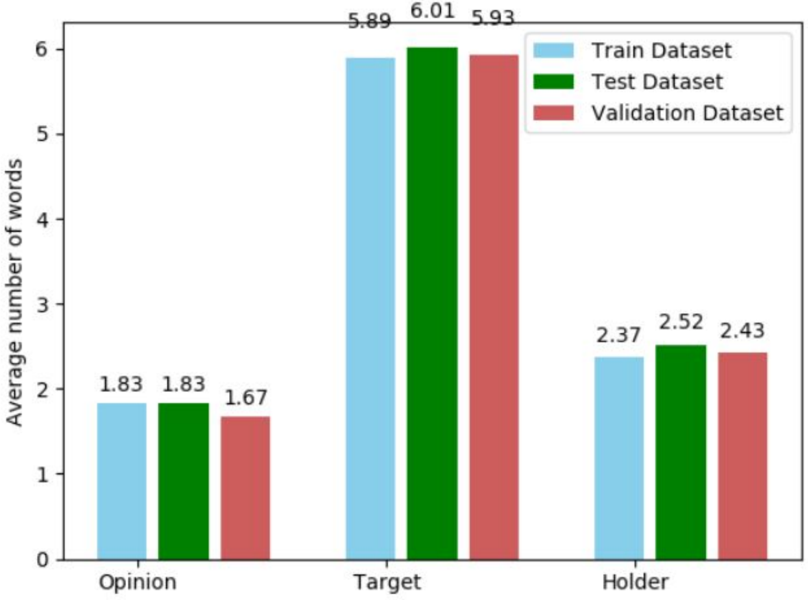


Figure 4.1.: Average number of words per opinion entity

Long target entities

- Lower precisions
- Lower recalls

for identifying target entities, especially when we evaluate the results using proportional overlap

Experiment: BiLSTM(+CRF)

- 3 layer BiLSTM
- Hidden units size = 50
- Kernel initializer: random uniform
- Dropout:
  - Fixed word2vec: 0.2
  - Others: 0.5
- Optimizer: adadelta
- Batch Size: 48
- **Embedding size: 300**
- CRF: marginal mode

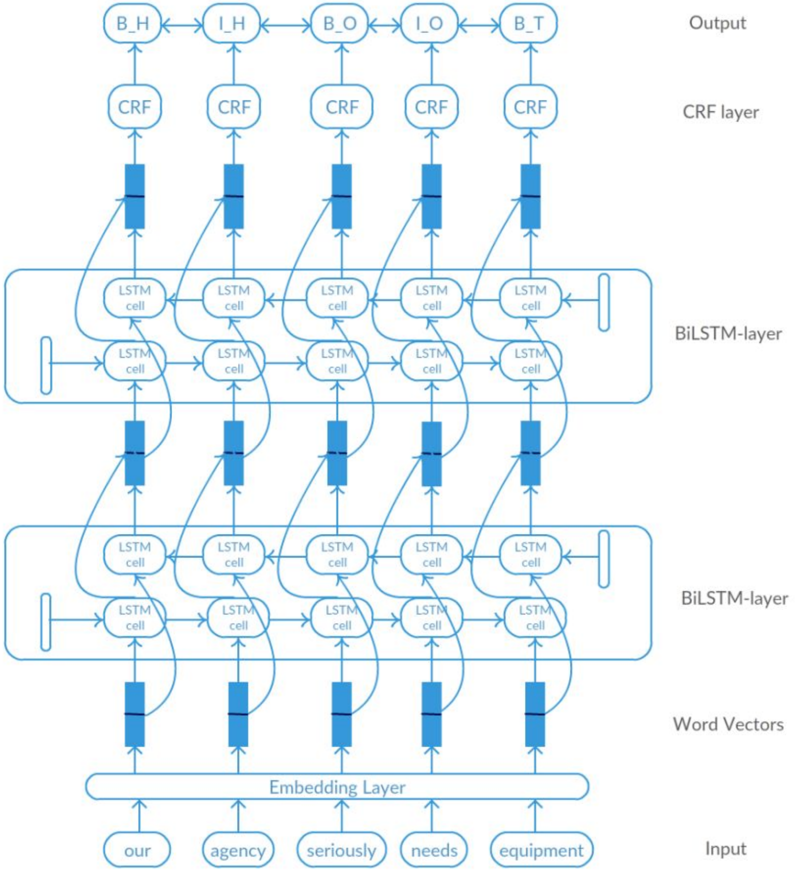
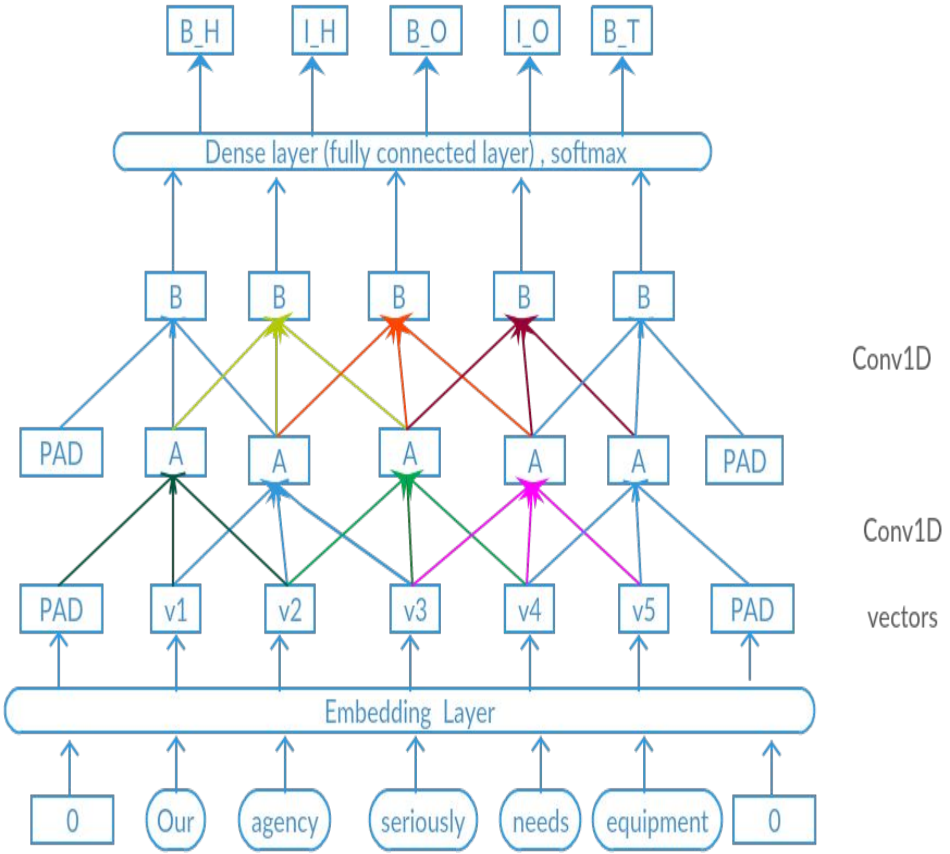


Figure 3.5.: 2-layer BiLSTM + CRF

Experiment: CNN(+CRF)

- 4 CNN layers
- Kernel size (3 or 5)
- Filters = 50
- Dropout rate = 0.5
- Optimizer : adadelta
- **Embedding size:**
  - Pre-trained word2vec embeddings: 300
  - Random embeddings: 100
- Weight initializer: Glorot Uniform
- Balanced sample weights
- CRF: marginal mode



Quantitative Analysis

- BiLSTM with trainable Word2Vec: strong baseline, best quantitative result.
- BiLSTM+CRF is less reliant on pre-trained embeddings, compared to BiLSTM.
- CNN(+CRF) models achieve better performance with trainable Word2Vec embeddings.
- CNN(+CRF) models have much higher recalls and lower precisions.

Binary Overlap		Opinion Expression			Opinion Target			Opinion Holder		
Method		Precision	Recall	F1 Score	Precision	Recall	F1 Score	Precision	Recall	F1 Score
BiLSTM	Word2Vec+trainable	75.96	**74.62**	**75.28**	26.67	36.38	30.78	67.67	**69.14**	**68.4**
	fixed Word2Vec	74.38	71.39	72.85	28.46	41.86	**33.88**	66.64	65.19	65.91
	random Embedding	67.10	67.33	67.21	17.9	41.86	25.08	60.51	56.59	58.48
BiLSTM+CRF	Word2Vec+trainable	**77.70**	72.27	74.89	30.29	29.98	30.13	**69.79**	66.77	68.25
	fixed Word2Vec	75.45	70.45	72.86	26.22	**45.89**	33.37	68.15	62.67	65.3
	random Embedding	71.39	71.39	71.39	25.9	37.66	30.69	63.27	66.61	64.9
CNN	Word2Vec+trainable	52.18	89.13	65.82	**20.15**	73.67	31.64	47.25	83.03	60.23
	fixed Word2Vec	42.65	**92.30**	58.34	15.21	**86.65**	25.88	43.74	83.74	57.46
	random Embedding	47.93	85.25	61.36	18.17	77.51	29.44	43.46	77.11	55.59
CNN+CRF	Word2Vec+trainable	**54.31**	87.37	**66.98**	19.62	84.28	**31.83**	**49.88**	78.77	**61.08**
	fixed Word2Vec	46.52	91.07	61.58	16.22	85.56	27.27	44.21	**85.0**	58.17
	random Embedding	51.71	80.26	62.9	18.27	77.15	29.54	45.94	73.01	56.39
Proportional Overlap		Opinion Expression			Opinion Target			Opinion Holder		
Method		Precision	Recall	F1 Score	Precision	Recall	F1 Score	Precision	Recall	F1 Score
BiLSTM	Word2Vec+trainable	71.09	**67.44**	**69.2**	23.56	29.91	26.36	65.18	**64.93**	**65.05**
	fixed Word2Vec	69.17	63.42	66.17	25.65	32.09	28.51	63.25	58.67	60.87
	random Embedding	61.81	59.16	60.46	15.53	32.75	21.07	57.62	51.37	54.32
BiLSTM+CRF	Word2Vec+trainable	**74.22**	63.76	68.59	**28.51**	22.26	25.0	**67.34**	62.85	65.02
	fixed Word2Vec	71.11	62.32	66.43	22.76	**36.23**	**27.96**	65.22	57.13	60.91
	random Embedding	64.31	65.17	64.74	18.16	33.19	23.48	58.43	63.62	60.91
CNN	Word2Vec+trainable	41.8	83.26	55.66	**18.15**	50.04	**26.64**	43.55	76.64	55.54
	fixed Word2Vec	32.68	**87.24**	47.55	12.41	69.04	21.04	38.06	76.63	50.86
	random Embedding	36.52	78.18	49.78	15.55	52.93	24.04	39.31	68.94	50.07
CNN+CRF	Word2Vec+trainable	**43.18**	81.43	**56.43**	16.05	64.83	25.73	**45.21**	72.62	**55.73**
	fixed Word2Vec	36.7	85.25	51.31	13.14	**69.33**	22.09	38.01	**78.03**	51.12
	random Embedding	41.18	72.73	52.59	14.76	55.17	23.29	41.59	65.01	50.73

Qualitative Analysis

Method	Sentence
S1	However, Chavez who is known for his ala Fidel Castro left-leaning anti-American philosophy had on a number of occasions rebuffed US requests for more oil exports.
BiLSTM +Word2Vec +trainable	However, Chavez who is known for his ala Fidel Castro left-leaning anti-American philosophy had on a number of occasions rebuffed US requests for more oil exports.
BiLSTM +fixed Word2Vec	However, Chavez who is known for his ala Fidel Castro left-leaning anti-American philosophy had on a number of occasions rebuffed US requests for more oil exports.
BiLSTM +random Embedding	However, Chavez who is known for his ala Fidel Castro left-leaning anti-American philosophy had on a number of occasions rebuffed US requests for more oil exports.
BiLSTM +CRF +Word2Vec +trainable	However, Chavez who is known for his ala Fidel Castro left-leaning anti-American philosophy had on a number of occasions rebuffed US requests for more oil exports.
BiLSTM +CRF +fixed Word2Vec	However, Chavez who is known for his ala Fidel Castro left-leaning anti-American philosophy had on a number of occasions rebuffed US requests for more oil exports.
BiLSTM +CRF +random Embedding	However, Chavez who is known for his ala Fidel Castro left-leaning anti-American philosophy had on a number of occasions rebuffed US requests for more oil exports.
CNN +Word2Vec +trainable	However , Chavez who is known for his ala Fidel Castro left-leaning anti-American philosophy had on a number of occasions rebuffed US requests for more oil exports .
CNN +fixed Word2Vec	However , Chavez who is known for his ala Fidel Castro left-leaning anti-American philosophy had on a number of occasions rebuffed US requests for more oil exports .
CNN +random Embedding	However , Chavez who is known for his ala Fidel Castro left-leaning anti-American philosophy had on a number of occasions rebuffed US requests for more oil exports .
CNN +CRF +Word2Vec +trainable	However , Chavez who is known for his ala Fidel Castro left-leaning anti-American philosophy had on a number of occasions rebuffed US requests for more oil exports .
CNN +CRF +fixed Word2Vec	However , Chavez who is known for his ala Fidel Castro left-leaning anti-American philosophy had on a number of occasions rebuffed US requests for more oil exports.
CNN +CRF +random Embedding	However , Chavez who is known for his ala Fidel Castro left-leaning anti-American philosophy had on a number of occasions rebuffed US requests for more oil exports .

BiLSTM(+CRF) perform better than CNN(+CRF).

- CNN(+CRF) models have a lot of predictions, but usually inaccurate. → high recall, low precision

- BiLSTM+CRF is less reliant on pre-trained embeddings than BiLSTM models.

Qualitative Analysis

S2	A short while ago, our correspondent in Bethlehem said that Ra'fat al-Bajjali was martyred of wounds he sustained in the explosion.
BiLSTM +Word2Vec +trainable	A short while ago, our correspondent in Bethlehem said that Ra'fat al-Bajjali was martyred of wounds he sustained in the explosion.
BiLSTM +fixed Word2Vec	A short while ago, our correspondent in Bethlehem said that Ra'fat al-Bajjali was martyred of wounds he sustained in the explosion.
BiLSTM +random Embedding	A short while ago, our correspondent in Bethlehem said that Ra'fat al-Bajjali was martyred of wounds he sustained in the explosion.
BiLSTM +CRF +Word2Vec +trainable	A short while ago, our correspondent in Bethlehem said that Ra'fat al-Bajjali was martyred of wounds he sustained in the explosion.
BiLSTM +CRF +fixed Word2Vec	A short while ago, our correspondent in Bethlehem said that Ra'fat al-Bajjali was martyred of wounds he sustained in the explosion.
BiLSTM +CRF +random Embedding	A short while ago, our correspondent in Bethlehem said that Ra'fat al-Bajjali was martyred of wounds he sustained in the explosion.
CNN +Word2Vec +trainable	A short while ago, our correspondent in Bethlehem said that Ra'fat al-Bajjali was martyred of wounds he sustained in the explosion.
CNN +fixed Word2Vec	A short while ago, our correspondent in Bethlehem said that Ra'fat al-Bajjali was martyred of wounds he sustained in the explosion.
CNN +random Embedding	A short while ago, our correspondent in Bethlehem said that Ra'fat al-Bajjali was martyred of wounds he sustained in the explosion.
CNN +CRF +Word2Vec +trainable	A short while ago, our correspondent in Bethlehem said that Ra'fat al-Bajjali was martyred of wounds he sustained in the explosion.
CNN +CRF +fixed Word2Vec	A short while ago, our correspondent in Bethlehem said that Ra'fat al-Bajjali was martyred of wounds he sustained in the explosion.
CNN +CRF +random Embedding	A short while ago, our correspondent in Bethlehem said that Ra'fat al-Bajjali was martyred of wounds he sustained in the explosion.

Table 5.2.: Output from different models with label annotations.

- CNN(+CRF) models can identify Out-of-Vocabulary (OOV) words (e.g. Ra'fat al-Bajjali), which are not in Word2Vec.
- BiLSTM(+CRF) models trained on random embeddings also identify OOV words.

Conclusion

- CNN(+CRF) models can capture out-of-vocabulary words, but the outcome is more likely fragmented for long entities.
- BiLSTM(+CRF) models are more suitable for sequence labeling task (better quantitative and qualitative results), and less reliant on pre-trained word-embeddings.

Future Work

- Using CNN to encode character-level information of a word to its character-level representation.
- Assign all out-of-vocabulary words untrained but unique random vectors, rather than a common UNK vector.
- Model the entities in overlapping relations.
    - Have the same sample in multiple times with different labels (reserve sub-relations)

Thank you for your attention and support !

Evaluation Metrics

(10) a. (relevant) However, Chavez who is known for his ala Fidel Castro left-leaning anti-American philosophy had on a number of occasions rebuffed US requests for more oil exports.

b. (retrieved) However, Chavez who is known for his ala fidel castroleft-leaning anti-american philosophy had on a number of occasions rebuffed US requests for more oil exports.

c. (overlapped) However, Chavez who is known for his ala Fidel Castro left-leaning anti-American philosophy had on a number of occasions rebuffed US requests for more oil experts.

Binary Overlap	Precision	Recall	F1 Score
Opinion	$\frac{2}{3} = 0.67$	$\frac{2}{3} = 0.67$	0.67
Target	$\frac{1}{2} = 0.5$	$\frac{1}{1} = 1$	0.67
Holder	$\frac{1}{2} = 0.5$	$\frac{1}{2} = 0.5$	0.5
Proportional Overlap	**Precision**	**Recall**	**F1 Score**
Opinion	$(\frac{2}{2} + 0 + \frac{1}{1} + 0)/3 = 0.67$	$(\frac{2}{3} + 0 + \frac{1}{1} + 0)/3 = 0.56$	0.61
Target	$(\frac{2}{2} + 0)/2 = 0.5$	$(\frac{2}{6} + 0)/1 = 0.33$	0.40
Holder	$(\frac{1}{1} + 0)/2 = 0.5$	$(\frac{1}{1} + 0)/2 = 0.5$	0.5

Table 4.4.: Evaluation results of (10)

Opinion Entity	relevant	retrieved	overlapped
Opinion	is known for (3) ala Fidel Castro left-leaning anti-American (5) rebuffed (1) Ø	is known (2) Ø rebuffed (1) requests (1)	is known (2) Ø rebuffed (1) Ø
target	US requests for more oil exports (6) Ø	oil exports (2) ala Fidel Castro left-leaning anti-American (5)	oil exports (2) Ø
holder	Chavez (1) his (1) Ø	Chavez (1) Ø US (1)	Chavez (1) Ø Ø

Table 4.3.: Evaluating (10). Each expression is an entity. The number in (*) is the number of tokens in an entity. Ø indicates no labeled entity.

BiLSTM with trainable word embeddings