

Report on relation extraction

Yiqing Hua

October 22, 2015

1 General Settings

The following results are on trained on the same hyperparameter settings. Currently they are using the logistic regression without any regularization. Some settings include:

- OCLASS weight:

	Target	Agent	DSE
weight	0.3	0.8	0.5

- Network Layers: 2
- Learning Rate of the Classifiers: 0.1
- Word Vector Dimension: 25
- Tested and Trained on Bishan's data and her datasplits.

2 Experiments

Variant 1 The training samples are picked according to the labels predicted. While training, the spans that have no overlap with any gold standard answers are not included. No backpropagation to the neural networks. The feature vectors are extracted from the last hidden layer of the neural network. Each span has two feature vectors, the average of the forward hidden layer vectors and the average of the backward hidden layer vectors.

The results on entity extractions are shown as follows.

And the Relation results:

	Target		Agent		DSE	
	Prop.	Bin.	Prop.	Bin.	Prop.	Bin.
P	0.278156	0.317073	0.575763	0.588529	0.347986	0.411924
R	0.46463	0.587744	0.512373	0.536866	0.542221	0.561512
F1	0.347986	0.411924	0.590153	0.65977	0.537502	0.576856

	P	R	F1
is from	0.300813	0.318052	0.309192
is about	0.289109	0.418338	0.34192

Variant 2 The training samples are picked according to the labels predicted. While training, the spans that have no overlap with any gold standard answers are not included. **There’s backpropagation to the neural networks.**

The feature vectors are extracted from the last hidden layer of the neural network. Each span has one feature vector, the concatenation of the forward hidden layer vectors from the first token and the backward hidden layer vector from last token.

The results on entity extractions are shown as follows.

And the Relation results:

	Target		Agent		DSE	
	Prop.	Bin.	Prop.	Bin.	Prop.	Bin.
P	0.283911	0.320681	0.591958	0.607143	0.524426	0.540486
R	0.478922	0.612813	0.518698	0.543779	0.556322	0.616092
F1	0.35649	0.421036	0.552912	0.573717	0.539903	0.575818

	P	R	F1
is from	0.312227	0.409742	0.354399
is about	0.312	0.446991	0.367491

Variant 3 The training samples are picked according to the **gold standard labels**. There’s backpropagation to the neural networks. **And the learning rate on the output layer of the neural network does not decay.**

The feature vectors are extracted from the last hidden layer of the neural network. Each span has one feature vector, the concatenation of the forward hidden layer vectors from the first token and the backward hidden layer vector from last token.

The results on entity extractions are shown as follows.

	Target		Agent		DSE	
	Prop.	Bin.	Prop.	Bin.	Prop.	Bin.
P	0.266633	0.303738	0.606677	0.621083	0.514708	0.533719
R	0.490078	0.62117	0.47435	0.497696	0.577337	0.641379
F1	0.345365	0.407982	0.532415	0.552585	0.544227	0.582617

And the Relation results:

	P	R	F1
is from	0.302752	0.378223	0.336306
is about	0.3	0.378223	0.334601

3 Some Conclusion

- According to my other experiments, the relation classifier simply predicts most of the span pairs as true pairs. So it may be possible that it's backpropagating useless errors to the neural net. But this also indicates that the recall we have right now is the **upper bound** of recall we can ever get.
- Since the classifier can not tell the neural net some spans are false spans, maybe it will not improve the entity extraction results as we expected.
- Because now we have more information to backpropagate to the last hidden layer, tuning the decay rate may also be effective but sometimes it can also lead to gradient explosion.

I'm also trying to use the library Bishan used for linear regression, which is liblinear. And I tried to tune the weight of the classifier so it will not only make trivial suggestions.

However, the training results of the classifier doesn't improve with more epochs with neural network. The accuracy is from 40% to 60%. I'm training on gold standard pairs, among which around 58% of the agent dse pairs are related, and around 55% of the target dse pairs are related.