OPINION MINING WITH DEEP RECURRENT NEURAL NETWORKS

Ozan Irsoy and Claire Cardie

CORNELL UNIVERSITY FOUNDED A.D. 1865

Introduction

Fine-grained opinion analysis aims to detect subjectivity (e.g. “hate”) and characterize

- Intensity (e.g. strong)
- Sentiment (e.g. negative)
- Opinion holder, target or topic
- …

Important for a variety of NLP tasks such as

- Opinion-oriented question answering
- Opinion summarization

Introduction

In this work, we focus on detecting
direct subjective expressions (DSEs) and
expressive subjective expressions (ESEs).

DSE: Explicit mentions of private states or speech events expressing private states

ESE: Expressions that indicate sentiment, emotion, etc. without explicitly conveying them.

Example

The committee, [as usual]$_{ESE}$, [has refused to make any statements]$_{DSE}$.

In BIO notation (where a token is the atomic unit):

The	committee	,	as	usual	,	has
O	O	O	B_ESE	I_ESE	O	B_DSE

refused	to	make	any	statements	.
I_DSE	I_DSE	I_DSE	I_DSE	I_DSE	O

Related Work

Most earlier work formulated as a token-level sequence-labeling problem.

- Conditional Random Field (CRF) approaches (Breck et al. 2007)
- Joint detection of opinion holders with CRFs (Choi et al. 2005)
- Reranking approaches over a sequence labeler (Johansson and Moschitti, 2010 & 2011)
- Semi Markov CRF (semiCRF) based approaches, which operate at the phrase level rather than token level (Yang and Cardie, 2012 & 2013)

Related Work

Success of CRF based approaches hinges critically on access to a good feature set, typically based on

- Constituency and dependency parse trees
- Manually crafted opinion lexicons
- Named entity taggers
- Other preprocessing components

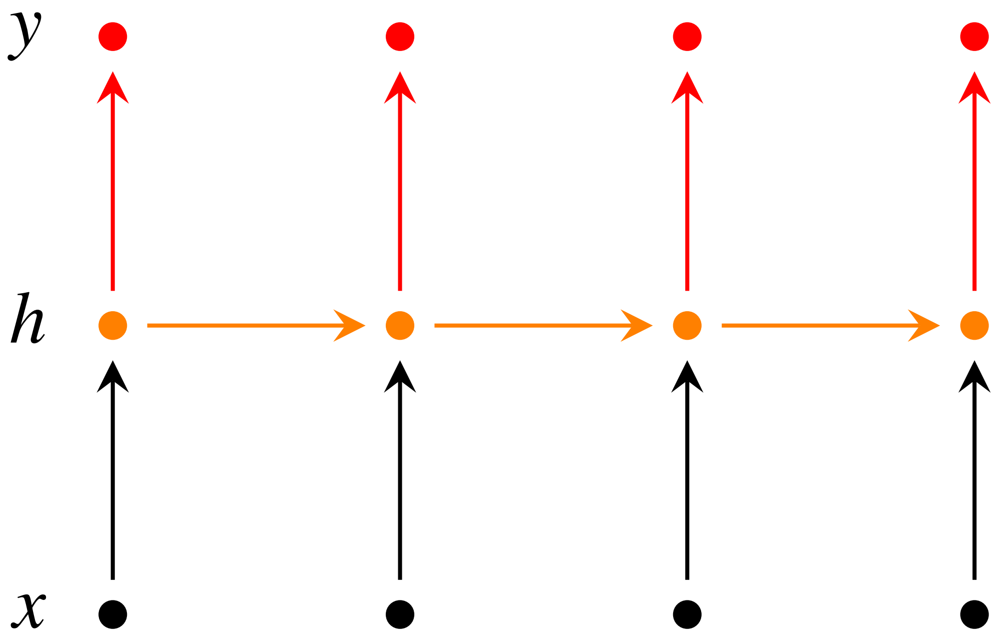
(See Yang and Cardie (2012) for an up-to-date list.)

What about feature learning?

Approach

- We adopt the same sequential prediction approach:
  A sentence is a sequence of tokens, each having a BIO based label.
- We use bidirectional shallow and deep Recurrent Neural Networks (RNN) for sequential prediction.
- RNNs have access to only a single feature set:
  Word vectors (which are trained in an unsupervised fashion).

Recurrent Neural Network



$$h_t = f(Wx_t + Vh_{t-1} + b)$$

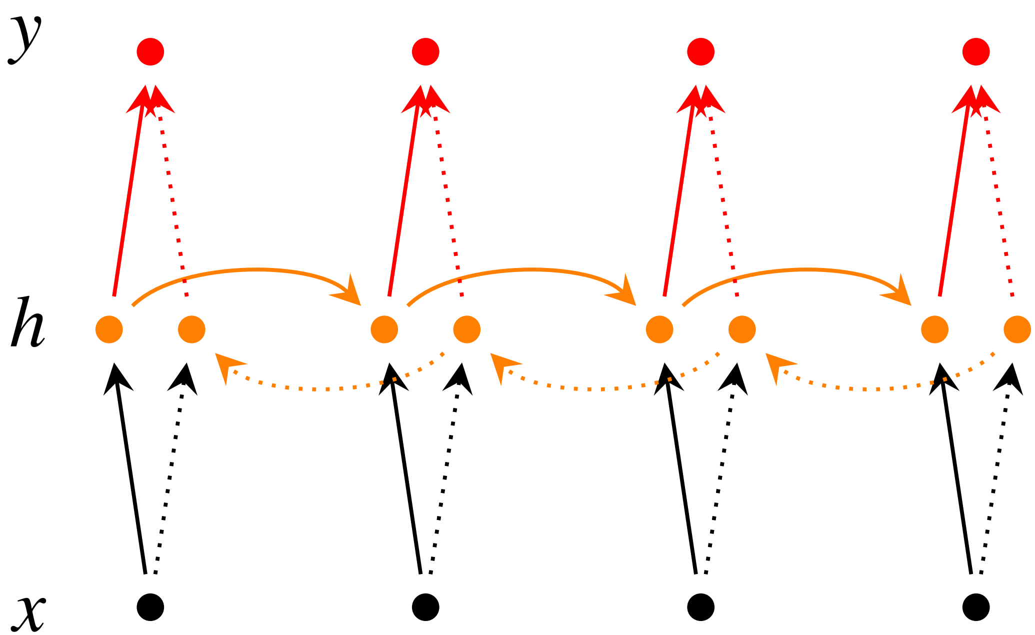
$$y_t = g(Uh_t + c)$$

x represents a token (word) as a vector.

y represents the output label (B, I or O).

h is the memory, computed from the past memory and current word. It summarizes the sentence up to that time.

Bidirectionality



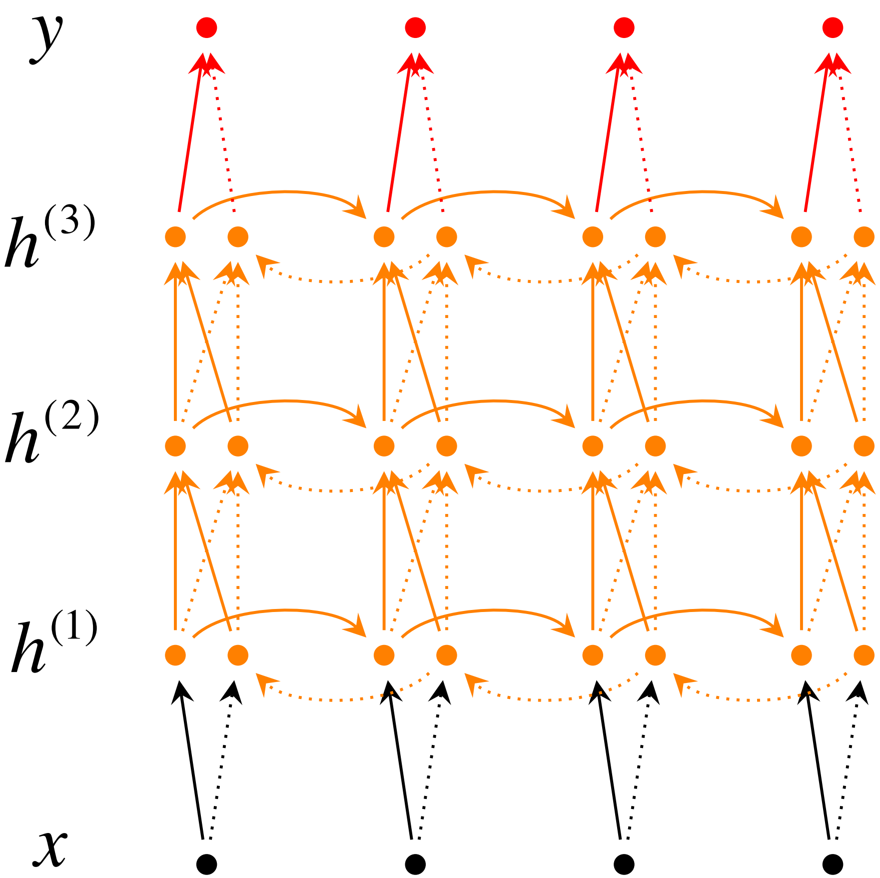
$$\overrightarrow{h}_t = f(\overrightarrow{W} x_t + \overrightarrow{V}\overrightarrow{h}_{t-1} + \overrightarrow{b})$$

$$\overleftarrow{h}_t = f(\overleftarrow{W} x_t + \overleftarrow{V}\overleftarrow{h}_{t+1} + \overleftarrow{b})$$

$$y_t = g(U[\overrightarrow{h}_t;\overleftarrow{h}_t] + c)$$

$h = [\overrightarrow{h};\overleftarrow{h}]$ now represents (summarizes) the past and future around a single token.

Going Deep



$$\overrightarrow{h}_t^{(i)} = f(\overrightarrow{W}^{(i)} h_t^{(i-1)} + \overrightarrow{V}^{(i)} \overrightarrow{h}_{t-1}^{(i)} + \overrightarrow{b}^{(i)})$$

$$\overleftarrow{h}_t^{(i)} = f(\overleftarrow{W}^{(i)} h_t^{(i-1)} + \overleftarrow{V}^{(i)} \overleftarrow{h}_{t+1}^{(i)} + \overleftarrow{b}^{(i)})$$

$$y_t = g(U[\overrightarrow{h}_t^{(L)}; \overleftarrow{h}_t^{(L)}] + c)$$

Each memory layer passes an intermediate sequential representation to the next.

Network Training

- Softmax and rectifier nonlinearities are used for output and hidden layer activations, respectively.
- Dropout regularization.
- Stochastic gradient descent with Cross-Entropy classification objective.
- Model selection is done via cross-validation over Proportional F1 metric.
- No pre-training, no fine-tuning.
- Two different parameter sizes: ~24k and ~200k. Therefore increasing depth cause a decrease in width.

Data

We use the MPQA 1.2 corpus (Wiebe et al., 2005) which consists of 535 news articles (11,111 sentences) that is manually labeled with DSE and ESEs at the phrase level.

As in previous work, we separate 135 documents as the development set to do model selection, and employ 10-fold cross-validation over the remaining 400 documents.

Performance Metrics

Exact boundaries are difficult, even for human annotators.

Two softer accuracy measures:

- Binary overlap: Every overlapping match between a predicted and true expression is correct.
- Proportional overlap: Every overlapping match is partially correct proportional to the overlapping amount (contribution of each match is in [0, 1]).

Binary and proportional Precision, Recall and F-measure are defined over these accuracy notions.

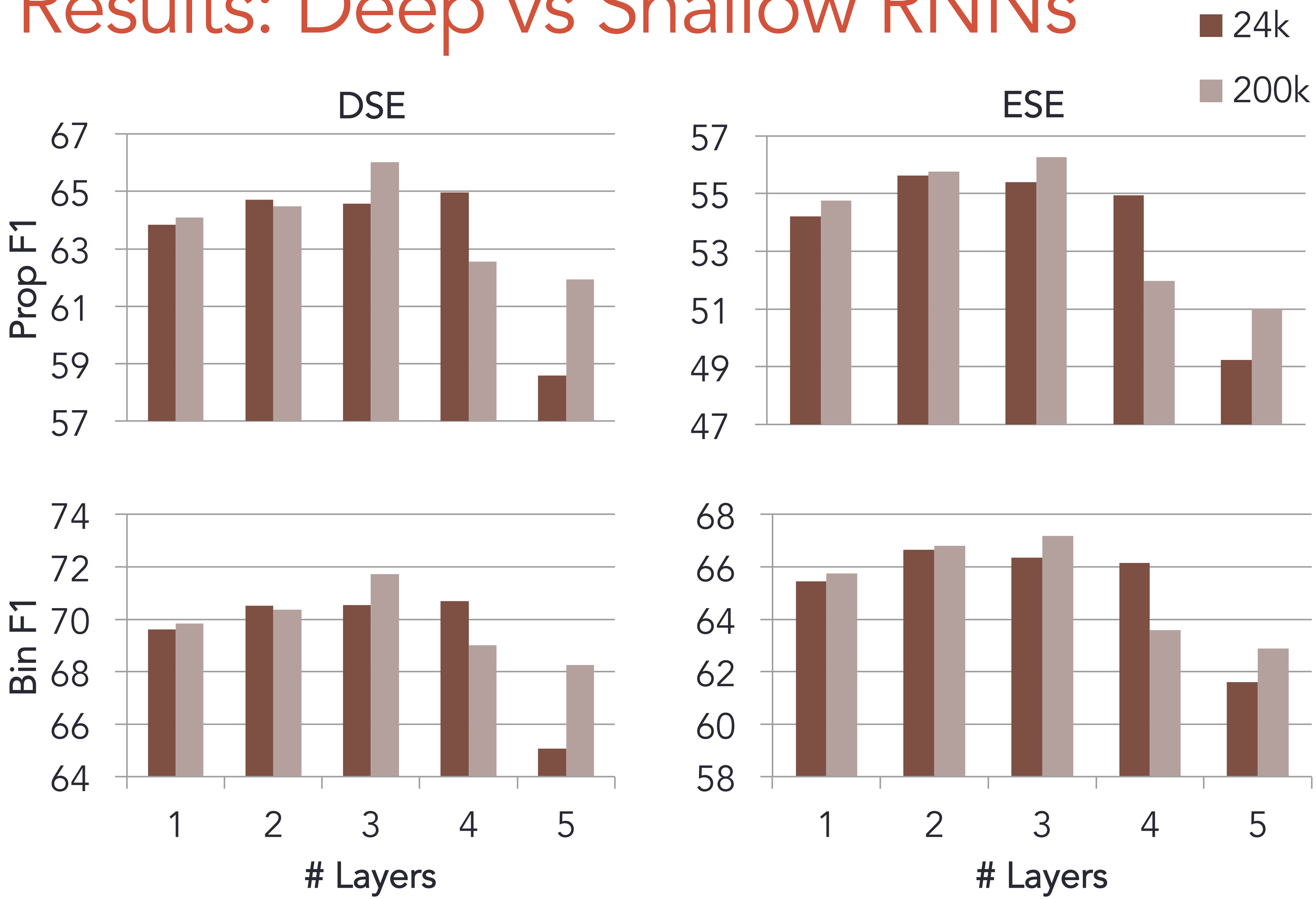
Hypotheses

We expected that deep RNNs would improve upon shallow RNNs, especially on ESE extraction.

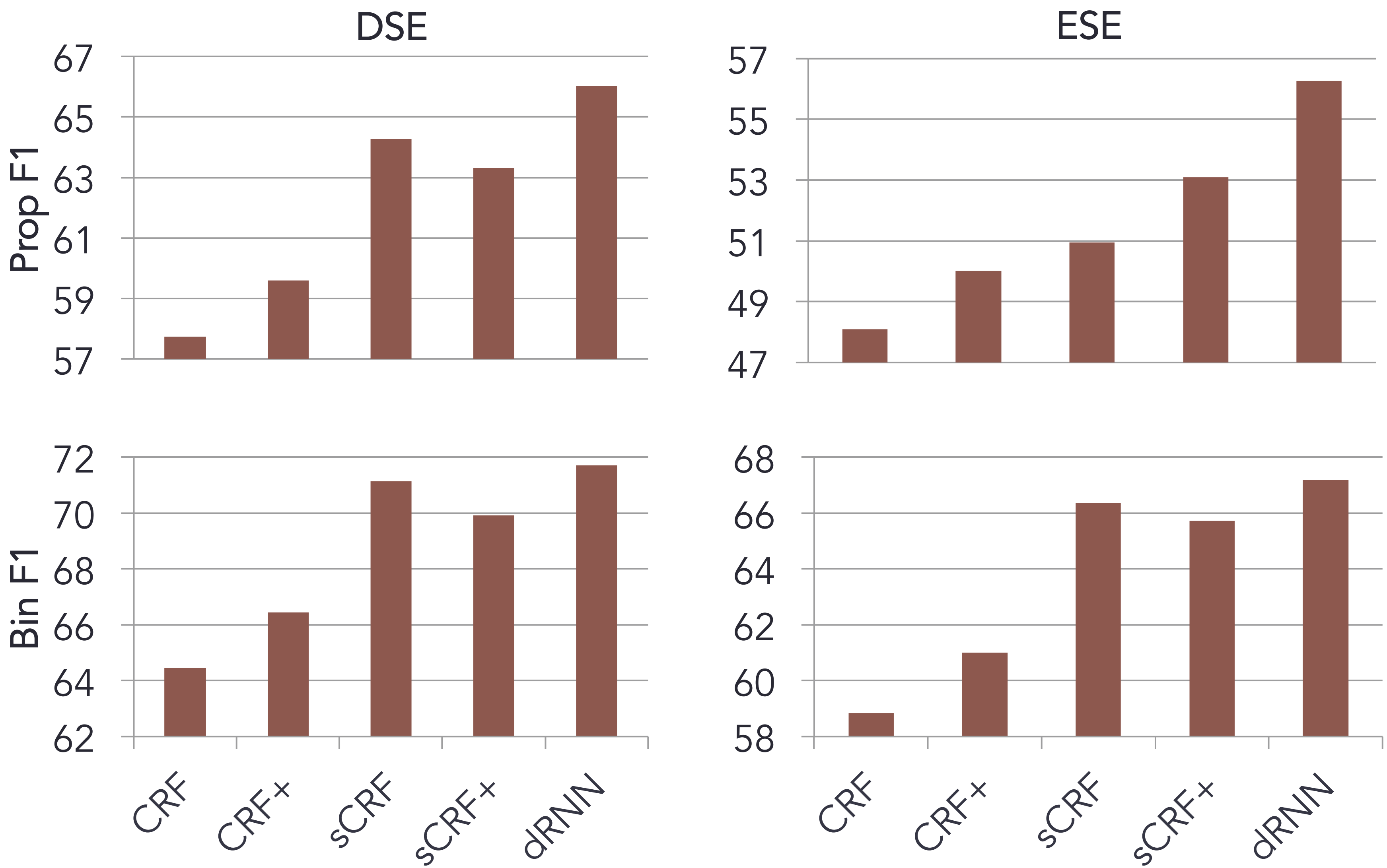
- ESEs are harder to identify: They are variable in length and might involve terms that are neutral in most contexts (e.g. “as usual”, “in fact”).

How RNNs would perform against (semi)CRFs was unclear, especially when CRFs are given access to word vectors.

Results: Deep vs Shallow RNNs



Results: DeepRNN vs (semi)CRF



Results: Examples

True	The situation obviously remains fluid from hour to hour but it [seems to be] [going in the right direction]
Deep RNN	The situation [obviously] remains fluid from hour to hour but it [seems to be going in the right] direction
Shallow RNN	The situation [obviously] remains fluid from hour to hour but it [seems to be going in] the right direction
Semi-CRF	The situation [obviously remains fluid from hour to hour but it seems to be going in the right direction]

Results: Examples

True	have always said this is a multi-faceted campaign [but equally] we have also said any future military action [would have to be based on evidence], …
Deep RNN	have always said this is a multi-faceted campaign but [equally we] have also said any future military action [would have to be based on evidence], …
Shallow RNN	have always said this is a multi-faceted [campaign but equally we] have also said any future military action would have to be based on evidence, …
Semi-CRF	have always said this is a multi-faceted campaign but equally we have also said any future military action would have to be based on evidence, …

Results: Examples

True	[In any case], [it is high time] that a social debate be organized …
Deep RNN	[In any case], it is [high time] that a social debate be organized …
Shallow RNN	In [any] case, it is high [time] that a social debate be organized …

True	Mr. Stoiber [has come a long way] from his refusal to [sacrifice himself] for the CDU in an election that [once looked impossible to win], …
Deep RNN	Mr. Stoiber [has come a long way from] his [refusal to sacrifice himself] for the CDU in an election that [once looked impossible to win], …
Shallow RNN	Mr. Stoiber has come [a long way from] his refusal to sacrifice himself for the CDU in an election that [once looked impossible] to win, …

Conclusion

- Deep recurrent nets perform better than their shallow counterparts of the same size on both DSE and ESE extraction.
- Both shallow and deep RNNs capture aspects of subjectivity, but deep RNNs seem to better handle the expression boundaries.
- Deep RNNs outperforms previous baselines CRF and semi-CRF without having access to the dependency or constituency trees, opinion lexicons or POS tags, even when (semi)CRF has access to word vectors.

Thanks!