Document-Level Entity-to-Entity Sentiment Analysis

Belinda Li

Introduction

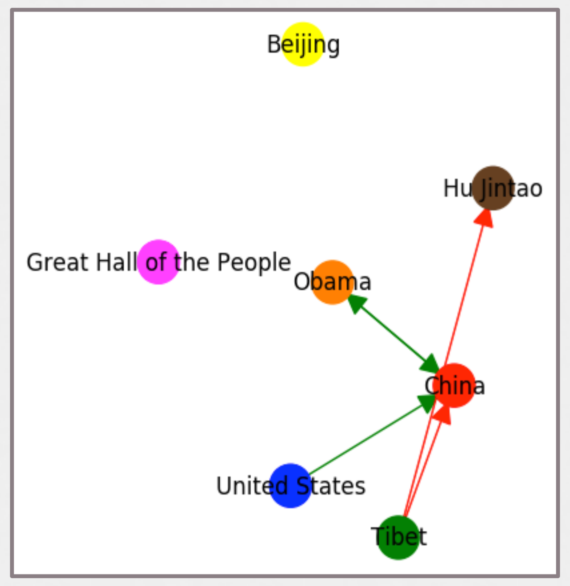
- News articles often encode complex networks of opinions amongst a multitude of entities.
- We aim to predict directed opinions (who feels positively towards whom) for all entities mentioned in an article

URGENT: Obama says U.S. recognizes Tibet as part of China.

United States President Barack Obama Tuesday said the U.S. government recognizes that Tibet is part of the People's Republic of China.

He also said that the United States supports the early resumption of dialogue between the Chinese government and representatives of the Dalai Lama to resolve any concerns and differences that the two sides may have.

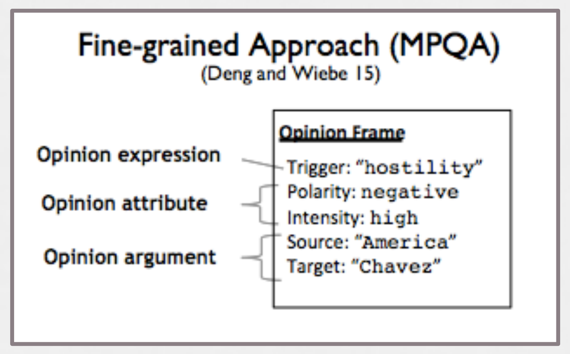
"The United States respects the sovereignty and territorial integrity of China," Obama said at a joint press conference with Chinese President Hu Jintao at Beijing's Great Hall of the People.



Related Work

Fine-grained sentiment analysis focuses on both the overall sentiment of a piece of text, as well as the entities involved (opinion holder/target)

- Most work in fine-grained sentiment analysis is at the sentence level
- Focuses on 1 sentiment relation between 1 entity pair, rather than interactions between many entities in a given document



Data

Used preprocessed datasets from Choi et al., 2016

1. train-original
2. dev-tune
3. dev-eval
4. test-KBP
5. test-MPQA

Creation of a new training set ("train-new"): Instead of using train-original, which has substantially different distribution from the development and test set, we create a new training dataset by combining the original training data with dev-tune and adding weakly generated "none" examples

- Weak "none" examples were generated by taking un-labeled pairs in the dataset and assuming they hold no sentiment.

Statistics

Dataset	Docs	Entities / Doc	Neg	None	Pos	% None
train-original	897	2.63	648	815	355	44.8
train-new	949	8.98	973	11306	1013	85.1
dev-tune	38	8.82	158	2349	365	81.8
dev-eval	37	9.08	174	2454	404	80.9
test-KBP	79	9.25	437	5459	718	82.5
test-MPQA	54	11.72	521	6464	625	84.9

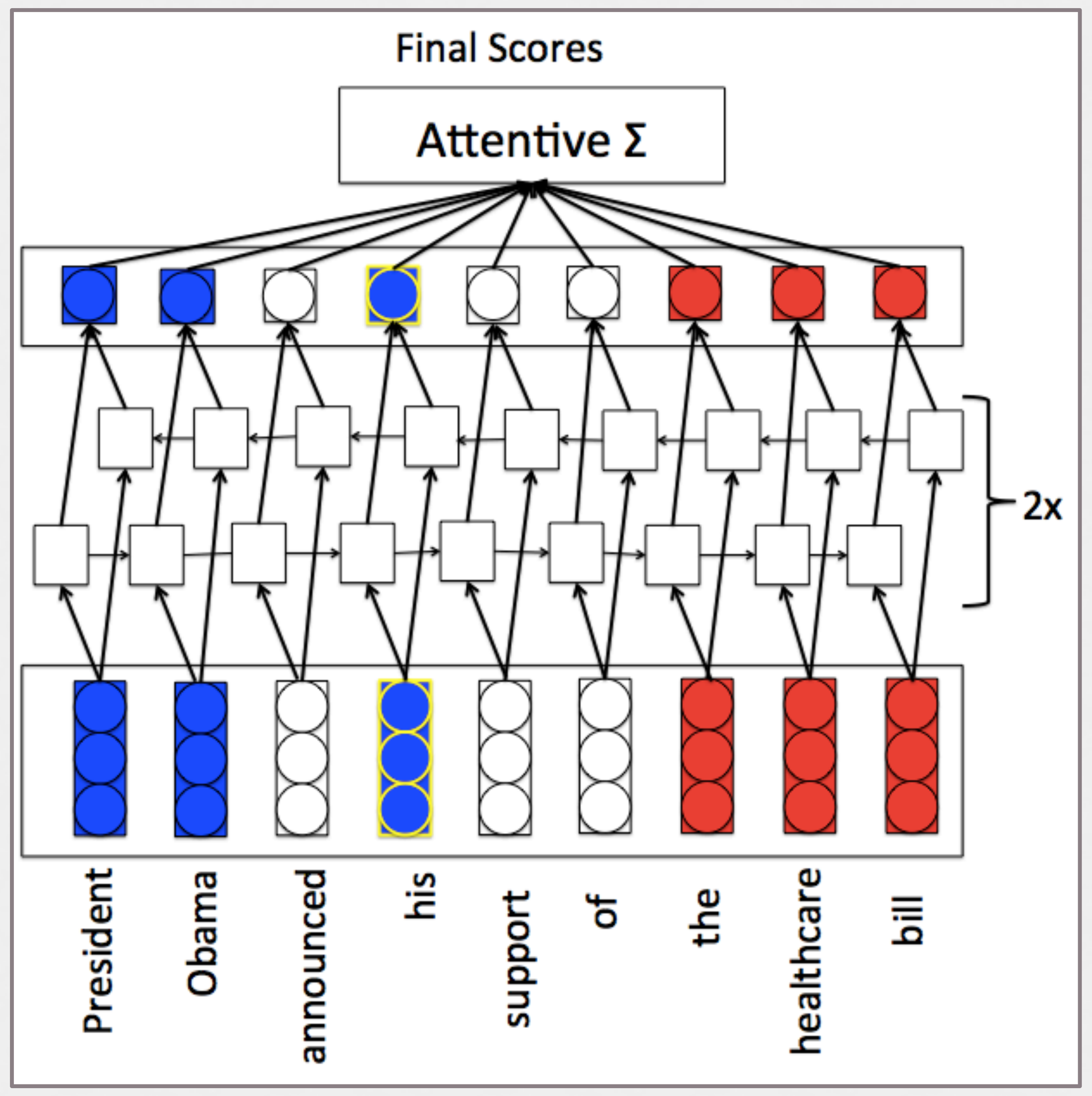
Models

Formulation of task: given a document and its E entities, look at all E(E-1) possible holder-target pairs and classify their sentiment relations into {POS, NEG, NONE}.

Sentence Baseline

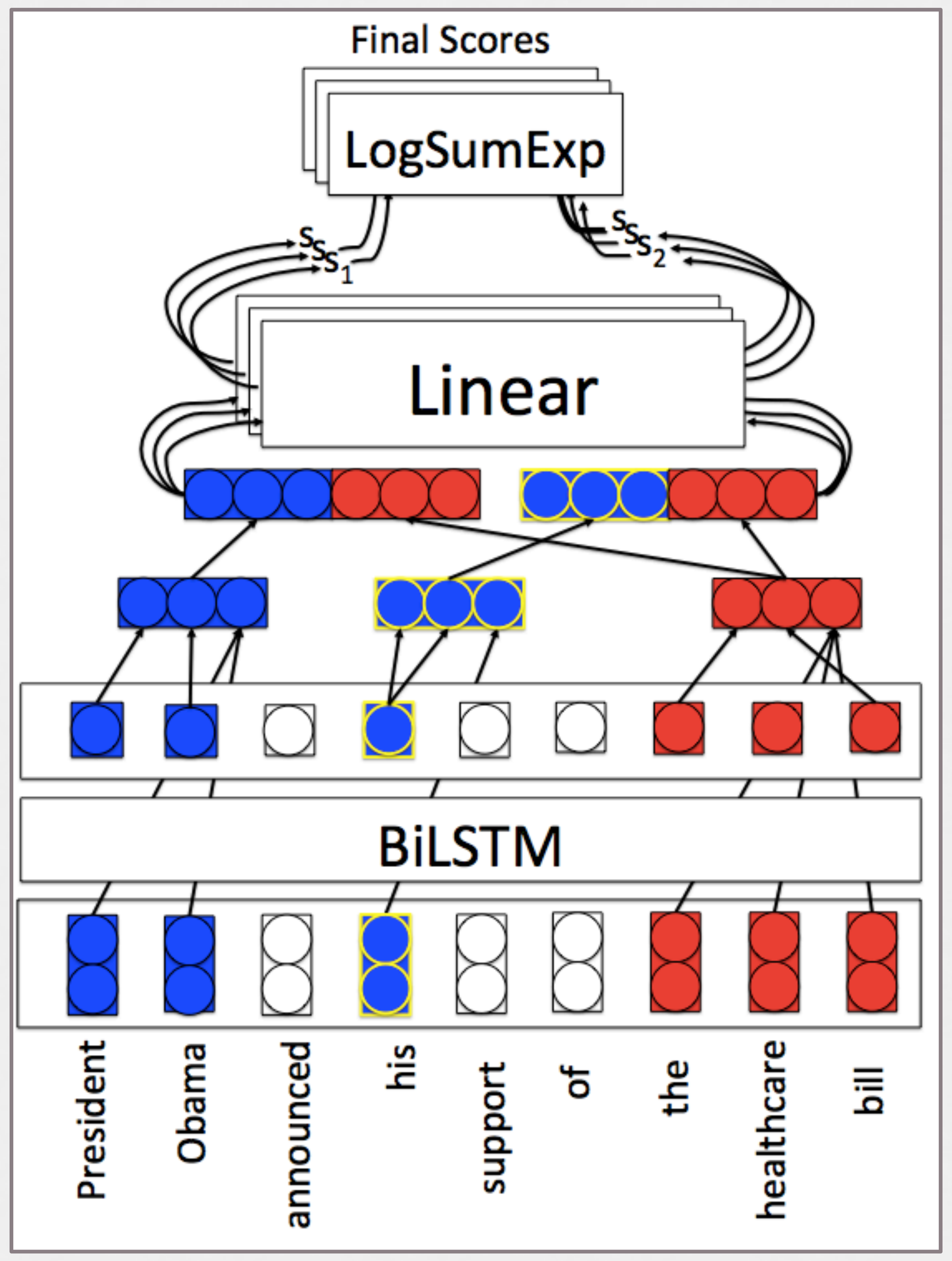
- Adapted from Socher et. al., 2013, a sentence-level neural model for extracting sentiment (of the text, in general), to perform entity-entity sentiment analysis at the document level
- Collect all sentences in which holder and target entity co-occur, and categorize their sentiment
- If # positive sentences > 0, classify pair as positive
- Else if # negative sentences > 0, classify pairs negative
- Otherwise classify as none

Attentive biLSTM Model



- Inputs: concatenation of pre-trained GloVe, and learned polarity and holder/target embeddings
- biLSTM: 2 layers (not shown)
- Attention: weights are calculated from a linear map of the LSTM-encoded tokens

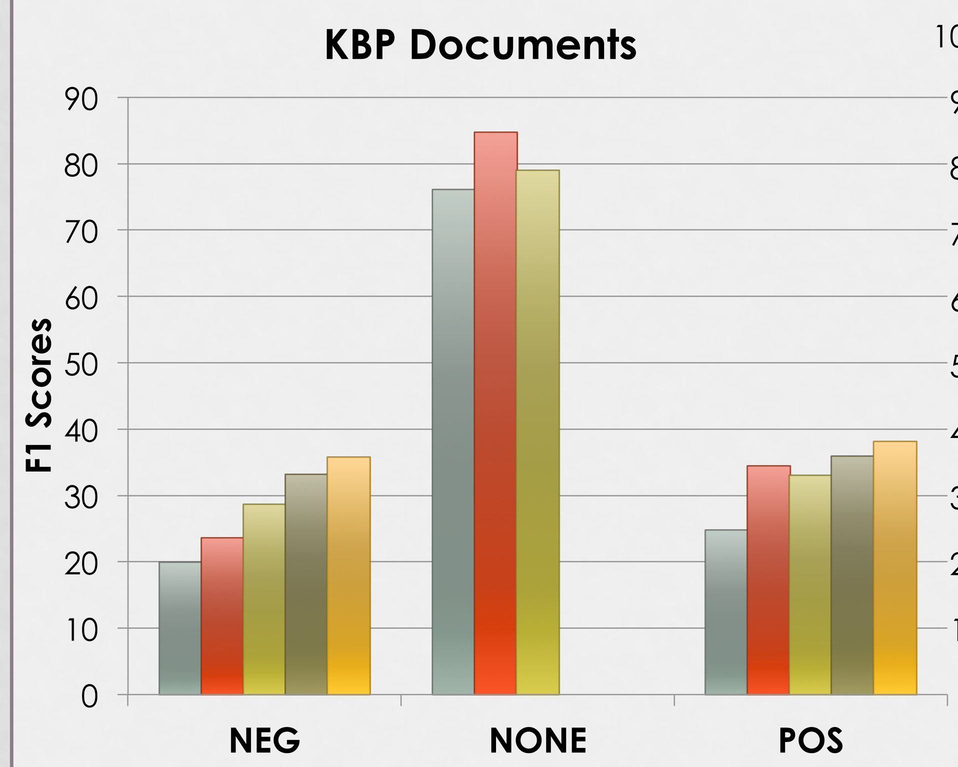
Pairwise Attentive biLSTM Model

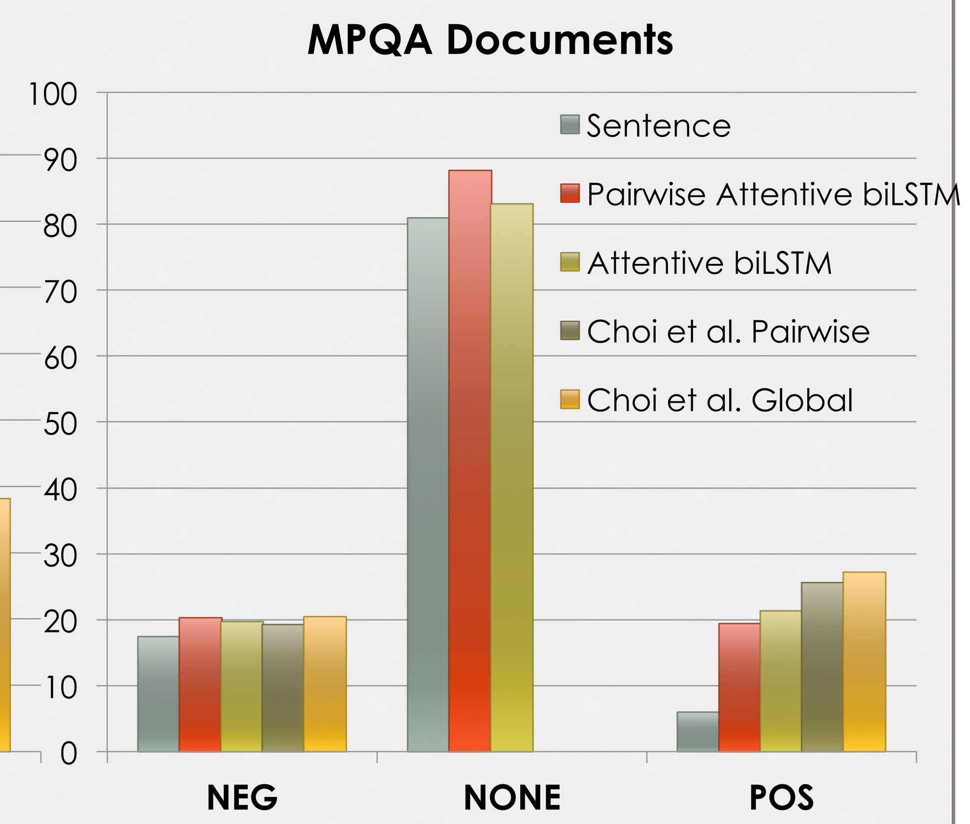


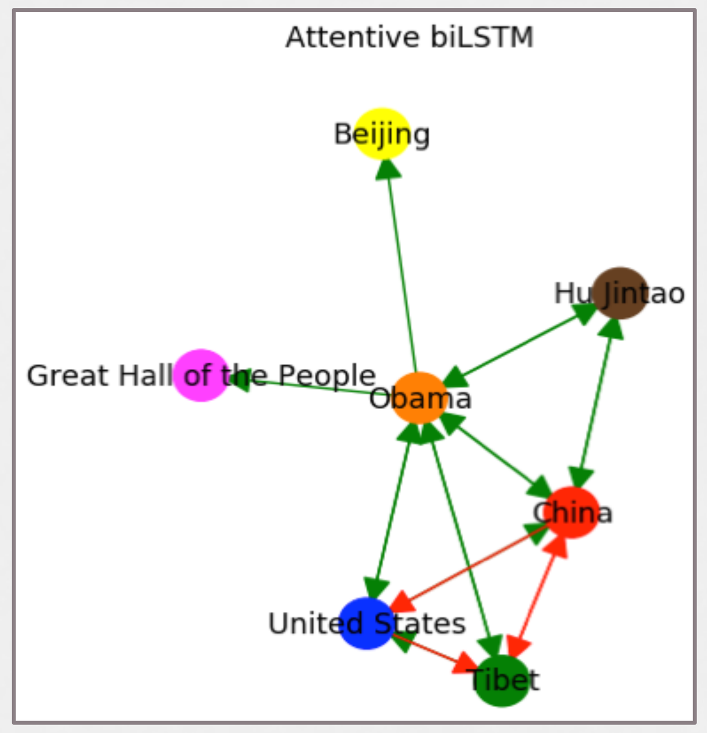
- Inputs: concatenation of pre-trained GloVe and learned polarity embeddings
- biLSTM: 2 layers
- Span representation: [encoded head, encoded tail, attention over all embedded tokens]
- Mention pair representation: concatenate all combinations of holders and target mentions. 4 features are also appended to each pair representation at this stage:
  - 1. Sentence
  - 2. Co-occurrence
  - 3. # mentions
  - 4. # mentions rank
- Final linear mapping generates scores for each mention pair, for each sentiment (POS/NEG/NONE)
- LogSumExp aggregates across mention pair score to generate a single score per sentiment for the entity pair

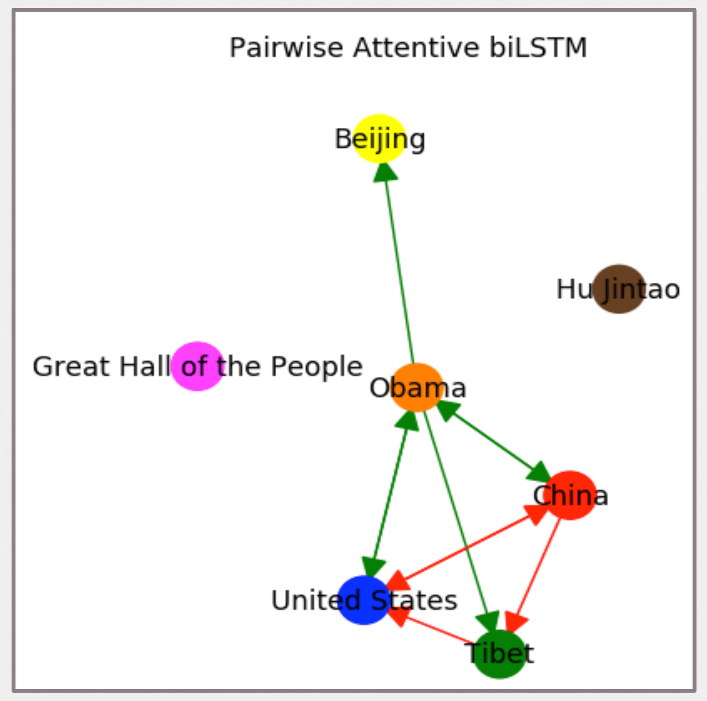
Results

- Precision, Recall, and F1-Scores for each label are reported

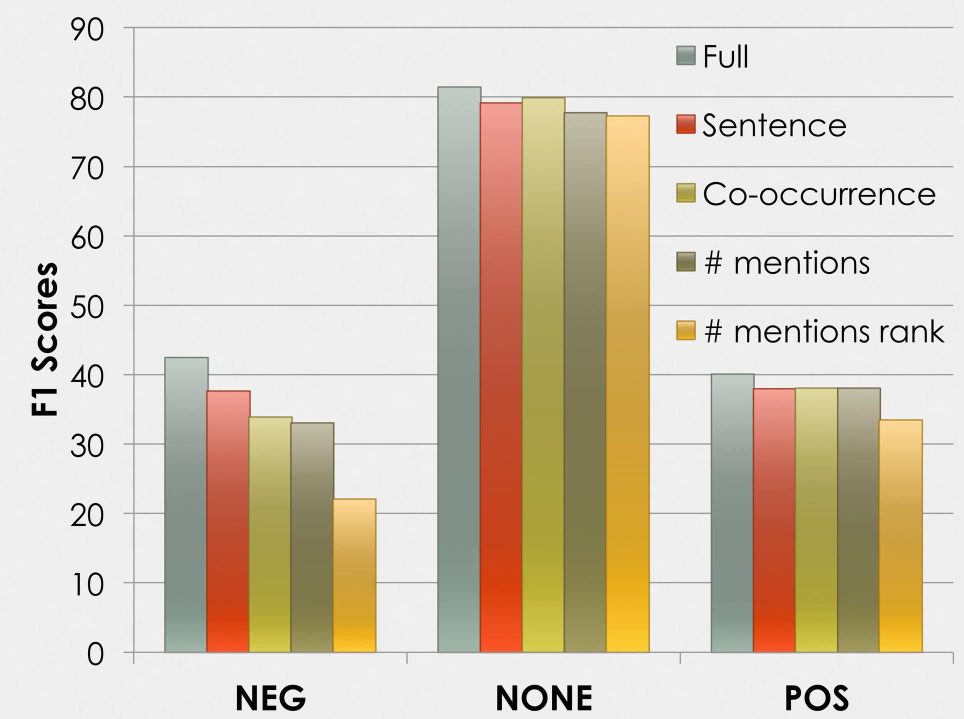








Ablations



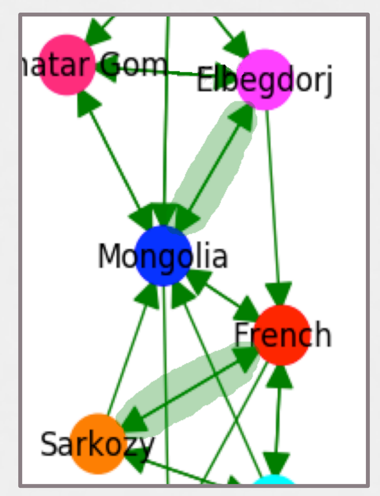
Discussion & Error Analysis

Sentence Baseline

- Sentiment of sentence =/= sentiment between entity pairs: Of the entities that co-occur once, the sentence that the entities co-occur in do not encapsulate the sentiment between the entities. This accounts for 73% of all errors

LSTM-based Models

- Overly dense annotations: Biggest category of errors are wrongly predicting "pos/neg" for "none" sentiment pairs.
- "None" sentiment is the most subjective & inconsistent annotations may make learning this category difficult for models
  - Should "<country> President <person>" be labeled positive?: This article (see diagram to right), containing "Mongolian President Tsakhia Elbegdorj…" and "French President Nicolas Sarkozy…" says yes. However, article in intro says no.



Conclusion

- Recurrent, LSTM-based neural models may not perform as well on this task as SVM models
- These models tend to be overeager in assigning sentiment
  - Might look into strategically pruning sentiment relations. One possibility is choosing the top-K most probable sentiment annotations per document
- Sentence-level models struggle to generalize to the document level
- May be helpful to only consider pairs of (holder, target) entities that make sense, i.e. only ever consider entities of type countries, persons, or organizations to be the holders
- Future work may look into encoding restrictions from social science theories into the network architecture or loss function