The SIGNLL Conference on Computational Natural Language Learning

CoNLL 2018

Author Response

Title: Who Blames Whom in a Crisis? Detecting Blame Ties from News Articles Using Neural Network
Authors: Shuailong Liang, Olivia Nicol and Yue Zhang

Instructions

The author response period has begun. The reviews for your submission are displayed on this page. If you want to respond to the points raised in the reviews, you may do so in the boxes provided below.

Please note: *you are not obligated to respond to the reviews*.

For reference, you may see the review form that reviewers used to evaluate your submission. If you do not see some of the filled-in fields in the reviews below, it means that they were intended to be seen only by the committee. See the review form HERE.

Review #1

Appropriateness (1-5):	5
Clarity (1-5):	3
Originality / Innovativeness (1-5):	3
Soundness / Correctness (1-5):	2
Meaningful Comparison (1-5):	4
Thoroughness (1-5):	4
Impact of Ideas or Results (1-5):	3
Recommendation (1-5):	2
Reviewer Confidence (1-5):	2
Recommendation for Best Paper Award (1-3):	1

Review

The paper introduces the task of Blame Tie Extraction: given a news article and a set of entities, determine the blame ties (A blames B) that exist between these entities.

It presents an approach for the task consisting of

- an Entity Prior Model, that learns blame source entity and target entity embeddings
- a Context Model, that learns contextual representations of entities involved in blame ties
- a joint model combining both

Experimental results on the blame tie extraction task are reported and analyzed.

Comments:

- Table 3: The target entities in the article and in the extracted blame ties disagree (see people (e_5), Hedge funds (e_4))

Comments on presentation:

- The paper could be revised for typos, especially with respect to the lack of articles and plurals.
- l. 299: W^e and b^e?

Strengths

- The introduction of the blame tie extraction task.
- The code will be released.

Weaknesses

- The dataset:
  - Its reliability, particularly of the test set, is questionable, since one of the authors, instead of multiple neutral subjects, created it and conducted the annotation of blame ties.
  - Data preparation for the context model: It relies on entity string matching, however, the position of the entities already have been provided by the created dataset.
- Experiments:
  - Evaluation on above dataset => reliable results?
  - No comparison to a baseline, such as the combination of two LSTMs (for blame source and target).
  - No comparison to related work, e.g., against approaches for relaction extraction.
  - Results: The model seems to overfit (Table 8), and the joint model, obtaining rather mixed results, is not convincing.
- Presentation of the approach: Details to the model and its design choices are not given or difficult to comprehend (see also the questions below and comments above).
- The task does not include the extraction of relevant entities, or at least the recognition of known entities beyond string matching, but could be modeled jointly with blame tie extraction. The inclusion of this subtask would strengthen the argument that blame tie detection systems can facilitate the work of social scientists.

Questions for Authors

- Model:

  Context model:

  - Why does Fig. 2 illustrate the model input without entity anonymization?
  - Why is entity string matching during training data creation required, instead of directly taking necessary information from the created dataset?

  Entity Prior Model: Could you please add an explicit explanation (Sect. 5.1) on how it is trained? (I.e., what do units 1 and 2 in the output layer stand for? Yes and no wrt to the existence of a blame tie?).
- Sect.3: What is the size of the list of blame and event related key words?
- Table 3: Why were no blame ties extracted for the last sentence in the article, involving bankers and hedge funds?

Review #2

Appropriateness (1-5):	5
Clarity (1-5):	3
Originality / Innovativeness (1-5):	3
Soundness / Correctness (1-5):	3
Meaningful Comparison (1-5):	4
Thoroughness (1-5):	3
Impact of Ideas or Results (1-5):	3
Recommendation (1-5):	3
Reviewer Confidence (1-5):	4
Recommendation for Best Paper Award (1-3):	1

Review

This paper proposes a new task (Blame Tie Extraction) with a new dataset based on the US financial crisis. The authors propose several neural baselines (one based on context-independent entity embeddings, one based on entity-anonymized context only, and one combining the two) and evaluate them on the task.

Strengths

- The authors provide an interesting new task and dataset (similar to relation extraction).
- The authors provide a thorough comparison of their neural models, including several hyperparameter variations, and they report results on both known and unknown entities (thus measuring generalization).

Weaknesses

- The paper doesn't compare the neural methods to any non-neural baselines. This is a major omission - in order to justify a neural solution, it's important to demonstrate that neural methods outperform simpler ones. Frequently, especially with these kinds of tasks, neural methods fail to improve upon simpler baselines.

I would like to see the results of the following non-neural baselines reported in the paper:

1. Random guessing. In Table 8, what F1 scores would random guessing get?
2. Simple rule-based system(s). For example, using a list of blame keywords, and labeling a blame relation whenever a keyword appears.
3. Simple statistical system for comparison with Entity Model. For example, in the KNOWN case, predict whether entity 1 blames entity 2 based on how often entity 1 blames entity 2 in the training data.
4. Simple statistical system for comparison with the Context Model. For example, you could use a simple ML model like Naive Bayes or SVM, with bag of word features.

5. Combination of 3 and 4 for comparison with the Combined Model.

    - The paper is sometimes hard to read or understand due to some unclear writing.

Questions for Authors

Thanks for committing to releasing your code and model! Please could you also report in the paper how long your models took to train (and on what hardware) - both number of training iterations and wall-clock time.

COMMENTS AND TYPOS:

The title is ungrammatical - "using neural networks" or "using a neural network" would be better.

From the abstract: "Using an end-to-end approach, we build a bi-directional Long Short-

Term Memory (LSTM) network on contexts where the entities appear to learn to automatically extract such blame ties on document level." This doesn't quite make sense. The entities don't learn to extract blame ties. The model learns to extract blame ties.

079: efforts -> effort 149: two folds -> two-fold 447: conduct develop experiments 495: internel 588: biLST 575: Table7 (important to fix this so that people can search for "Table 7") 577: vresion, does not improves

There are quite a few more typos and ungrammatical sentences.

Review #3

Appropriateness (1-5):	5
Clarity (1-5):	4
Originality / Innovativeness (1-5):	3
Soundness / Correctness (1-5):	3
Meaningful Comparison (1-5):	4
Thoroughness (1-5):	4
Impact of Ideas or Results (1-5):	3
Recommendation (1-5):	3
Reviewer Confidence (1-5):	4
Recommendation for Best Paper Award (1-3):	1

Review

The authors present a new dataset for "blame detection" in news articles, where the goal is to detect whether one entity "blames" another. The authors also present an LSTM+embedding approach to this task, achieving reasonable scores on their new dataset.

Overall, the paper introduces a creative task and is reasonably well-written. The embedding-based model that takes into account both prior entity relationships as well as linguistic context is very suitable for this task, and I appreciated the ablations and case-studies that teased apart the contributions of the different parts of the model.

However, there are several areas where the paper could be improved:

- Most prominently, the paper is lacking a non-neural baseline, so it is difficult to judge the performance of the presented models. For example, presenting the results from a simple bag-of-words random forest or gradient-boosted tree would greatly strengthen the paper. Of course, it is not entirely straightforward to phrase this task in a way that can be fed into a simple classifier, but there are several reasonable options (e.g., just ignore the entities and treat it as a binary classification task of "does X blame Y based on this linguistic context", as the context model does). Adding at least one non-trivial, non-neural baseline would greatly strengthen the paper.

- Another major issue with the paper is that it does not adequately explain the coding/annotating process. It appears that the authors themselves performed the annotation, but was there any check for inter-annotator agreement, etc? Adding some more detail about the annotation process and quality controls would strengthen the paper and likely increase the interest in the dataset.

- This work is quite related to stance detection (see, e.g., http://www.aclweb.org/anthology/S16-1003 for a description of this task), and the paper could benefit from a discussion of this related work.

- It seems that the entity embeddings will not be useful for generalization. This is implicit in Section 7.2 where the Entity Prior Model is simply ignored, but a more up-front discussion of this limitation and how to potentially get around it would be useful.

Strengths

- Introduces a novel, creative task.
- Develops a well-designed model for the novel task.
- Includes some thoughtful ablations and case-studies.

Weaknesses

- The paper is lacking a proper non-neural baseline.
- The dataset construction is not adequately described.
- The paper is missing a discussion of related work on stance detection.

Submit Response to Reviewers

Use the following boxes to enter your response to the reviews. Please limit the total amount of words in your comments to 900 words (longer responses will not be accepted by the system).

Response to Review #1:

Response to Review #2:

Response to Review #3:

General Response to Reviewers:

Response to Chairs

Use this textbox to contact the chairs directly only when there are serious issues regarding the reviews. Such issues can include reviewers who grossly misunderstood the submission, or have made unfair comparisons or requests in their reviews. Most submissions should not need to use this facility.

Submit