

# The Cornpittmich Chinese System for BeSt Evaluation 2016

Kai Sun, Xilun Chen, Yao Cheng, Xinya Du, and Claire Cardie  
Cornell University

# Overall Approach

- For target
  - Separate components for belief and sentiment
  - Each is a hybrid system
  - Rule-based + Machine learning-based
- For source
  - Genre-specific components for both belief and sentiment
  - Rule-based for both DF and NW

Belief

# Source: Rule-based

- Given a target candidate with its mention text/trigger,
  - For DF, its post author is the source
  - For NW, if there is a nearby word or phrase denoting reported speech (such as “说” (“say”), “指出” (“point out”)), regard the associated agent and the author of the article as the sources. Otherwise, regard the author of the article as the source

# Target: Hybrid

- Rule-based model
  - For DF
    - Always output type="cb" and polarity="pos" for each relation and event
  - For NW
    - Output type="cb" and polarity="pos" if the relation/event has only one source, or the source is not the article author
    - Output type="rob" and polarity="pos" if the relation/event has two sources, and the source is the article author
- A linear model\* for filtering
  - Take in the text around the relation/event mention and decide whether there is a belief or not. If the answer is no, it removes the corresponding belief output by the rule-based model from the final output

\*We used TextGrocery: <https://github.com/2shou/TextGrocery>

# Submissions

- DF: Rule + Linear
- NW: Rule\*

	System	Precision	Recall	F-score
DF	Baseline	0.808	<b>0.877</b>	<b>0.841+</b>
	Sys1,2,3	<b>0.839</b>	0.842	0.841-
NW	Baseline	<b>0.820</b>	0.602	<b>0.694</b>
	Sys1,2,3	0.583	<b>0.609</b>	0.596

Gold ERE, Test

\*Linear model was not used because we had no training data for NW

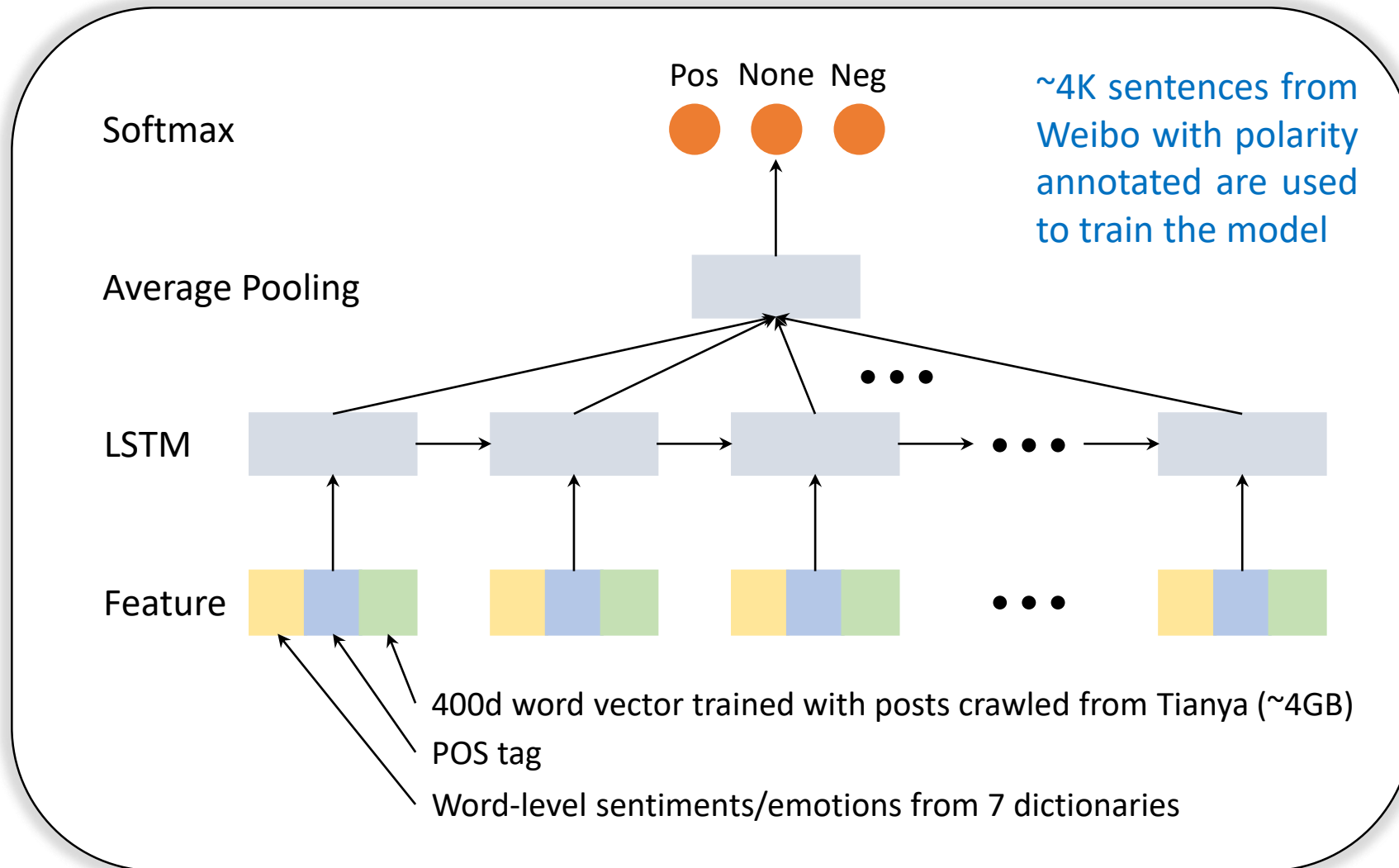
# Sentiment

# Source: Rule-based

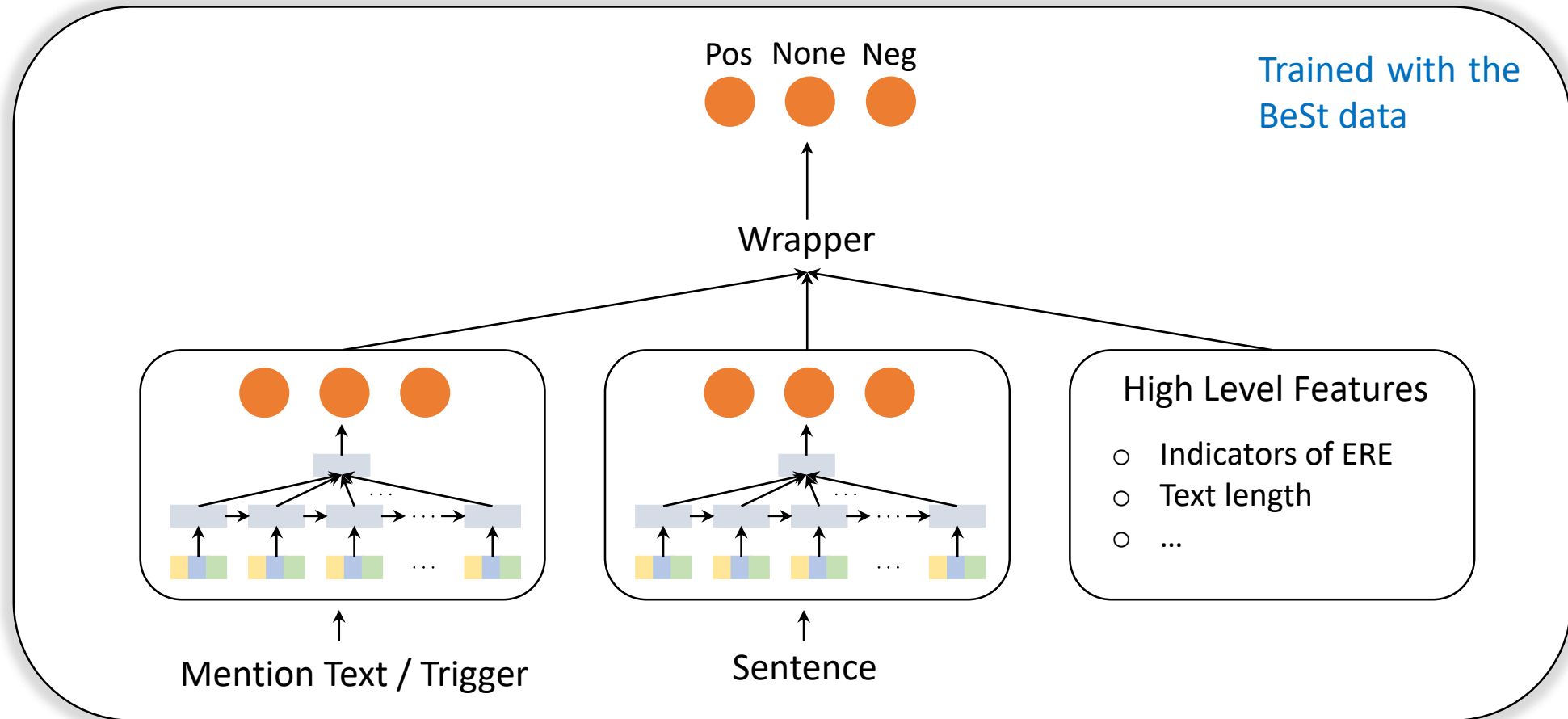
- Same as belief



# Target: Hybrid Sentence-level Model



# Target: Hybrid Model for BeSt



# Target: Hybrid Wrapper

- A set of data-driven rules with the goal of
  - Taking advantage of high-level features
  - Resolving inconsistent predictions from the mention text and the sentence
  - Setting different acceptance thresholds for different scenarios
- Examples
  - Different thresholds should be set for different types of target

# Target: Hybrid Wrapper

- A set of data-driven rules with the goal of
  - Taking advantage of high-level features
  - Resolving inconsistent predictions from the mention text and the sentence
  - Setting different acceptance thresholds for different scenarios
- Examples
  - Thresholds should be relaxed when the sentence the target entity belongs to has only one entity

# Target: Hybrid Wrapper

- A set of data-driven rules with the goal of
  - Taking advantage of high-level features
  - Resolving inconsistent predictions from the mention text and the sentence
  - Setting different acceptance thresholds for different scenarios
- Examples
  - When the mention text contains words with strong intensity, predictions at the sentence level should be discounted
    - 把枉法裁判、胡作非为、违法乱纪的腐败分子惩处工作抓好
    - Make punishing **corruption and corrupt elements** a success

# Submissions

- We use different  $F_\beta$ -score as the criteria for wrapper training

$$F_\beta = \frac{(1 + \beta^2) \cdot P \cdot R}{\beta^2 \cdot P + R}$$

- DF

$$\beta^2 = 1, 2.5, 0.2$$

- NW

$$\beta^2 = 2.5, 10, 1$$

	System	Precision	Recall	F-score
DF	Baseline	0.058	<b>0.771</b>	0.108
	Sys1	0.583	0.303	<b>0.399</b>
	Sys2	0.451	0.341	0.388
	Sys3	<b>0.600</b>	0.297	0.397
NW	Baseline	0.011	<b>0.340</b>	0.021
	Sys1	0.264	0.052	0.087
	Sys2	0.082	0.115	<b>0.096</b>
	Sys3	<b>0.298</b>	0.038	0.068

Gold ERE, Test

(Possibly) Interesting Observations for  
Sentiment





# Conclusion

- The task is challenging given limited number of annotations
- Our hybrid models have relatively good performance by taking advantage of human knowledge (in the hand-crafted rules), internal and external datasets.

# Conclusion

- The task is challenging given limited number of annotations
- Our hybrid models have relatively good performance by taking advantage of human knowledge (in the hand-crafted rules), internal and external datasets.

Thanks 😊

Any questions?