

Scalar Event Factuality

What did the author claim happened or will happen?

Aggregated Annotated Examples

U.S. embassies and military installations around the world were **ordered**^(3.0) to **set**^(2.6) up barriers and **tighten**^(2.6) security to **prevent**^(1.8) easy **access**^(-2.4) by unauthorized people.

Intel's most powerful computer chip has flaws that could **delay**^(0.8) several computer makers' marketing **efforts**^(2.6), but the "bugs" aren't **expected**^(-2.6) to **hurt**^(-2.0) Intel.

President Bush on Tuesday **said**^(3.0) the United States may **extend**^(1.6) its naval **quarantine**^(2.6) to Jordan's Red Sea port of Aqaba to **shut**^(1.4) off Iraq's last unhindered trade route.

He also **said**^(3.0) of **trade**^(-0.8) with Iraq: "There are no shipments at the moment."

Factuality is graded

Non-Expert Annotation

- Annotate all events in a binary task
- Brief instructions followed by 25 examples

We consider events to be things that may or may not occur either in the past, present or future (e.g., earthquake, meeting, jumping, talking, etc.). In some cases, it is not so clear whether a word is referring to an event or not. Consider these harder cases to be events.

And that, economists and investors contend, is why Germany and its Dutch stalking horse, Jeroen Dijsselbloem, the president of the Eurogroup of **finance** ministers, were so adamant that depositors - large and small, Cypriot and Russian - contribute 5.8 billion euros (\$ 7.5 billion) toward the 10 billion euro bailout of Cyprus's largest banks.

Does the highlighted word ('finance') refer to an event?

- Yes
 No

- Given annotated events, label events on a [-3,3] scale
- Brief instructions followed by 17 examples

On a scale from 3 to -3, rate how likely the highlighted event did or will happen according to the author of the sentence.

Ironically, the city is accused of **violating** an anti-corruption law that grew out of aborted plans to hold another GOP convention in San Diego a quarter-century ago.

On a scale from 3 to -3, how certain is the author that the highlighted event: 'violating' did or will happen? (required)

- 3: The author believes that it **certainly** did or will happen.
 2
 1
 0: The author's stance is **neutral**.
 -1
 -2
 -3: The author believes that it **certainly** did **not** or will **not** happen.

- Crowdsource events and their scalar factuality
- Annotate TempEval-3 corpus (superset of FactBank) for comparison.

Data	Documents	Sentences	Tokens
Train	194	2909	73220
Dev	64	1060	26146
Test	20	274	7004

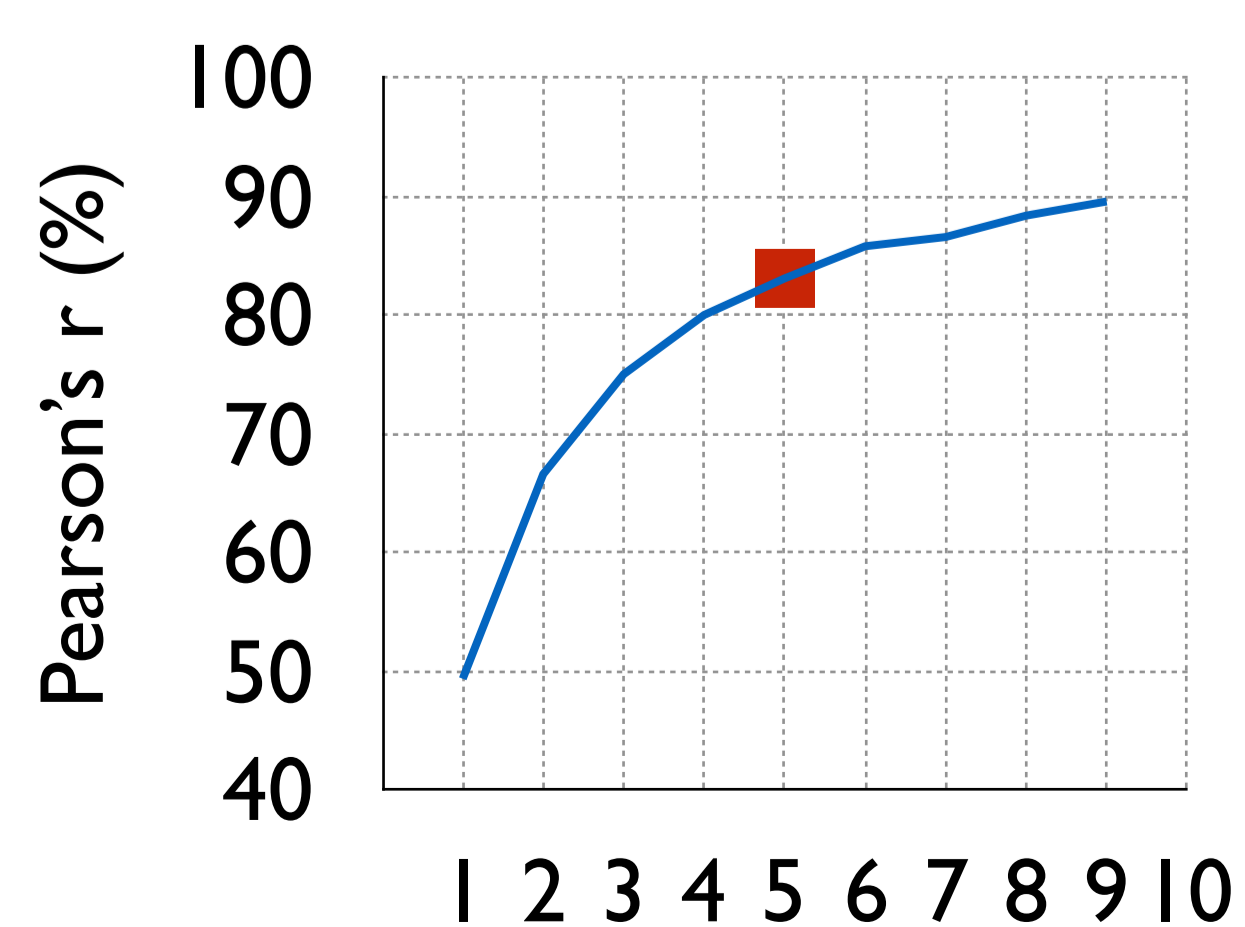
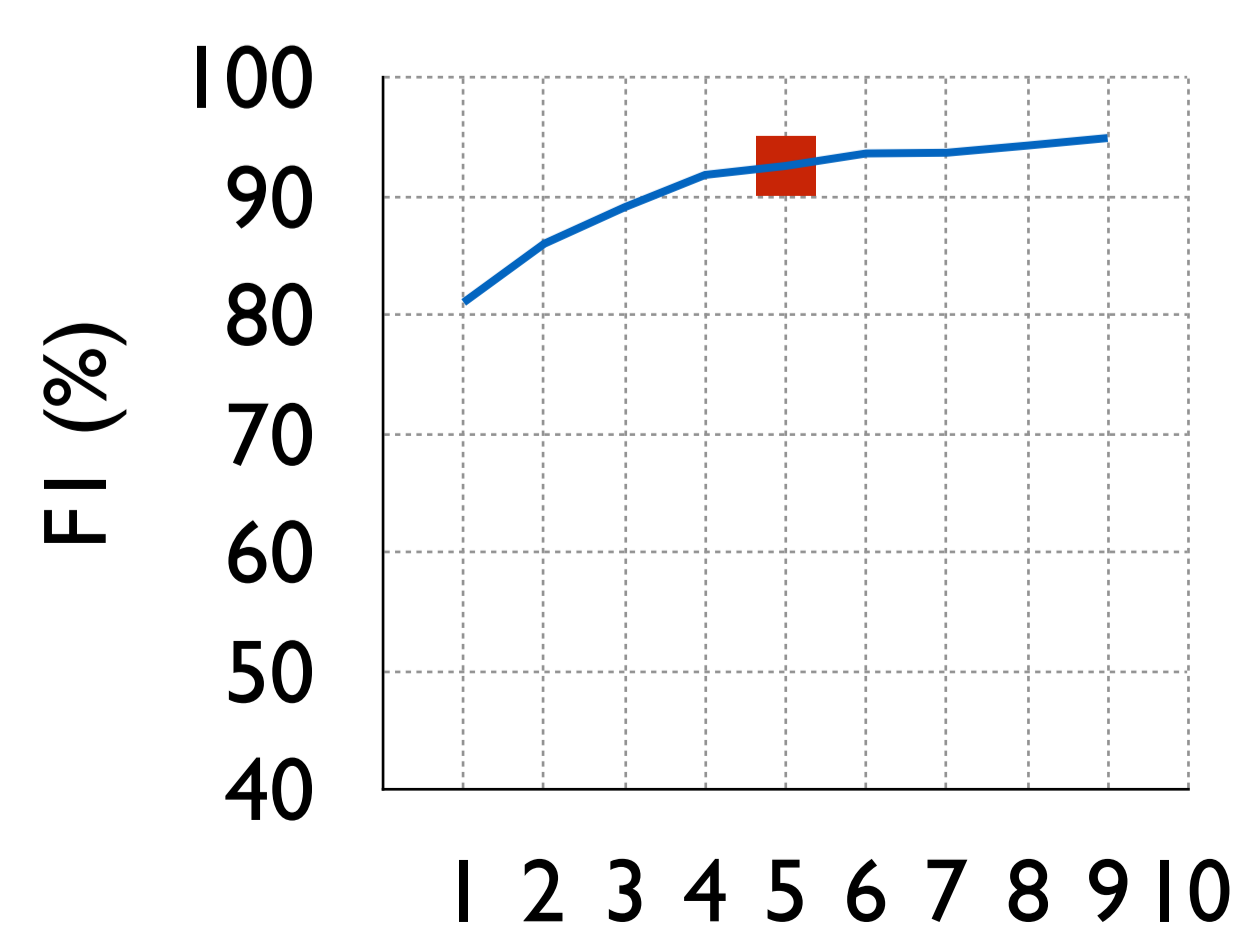
Discretized ratings	FactBank Labels									
	CT-	PR-	PS-	PS+	PR+	CT+	CTu	NA	Uu	
-3	39	0	0	0	0	0	0	0	29	
-2	29	2	0	0	0	0	0	0	44	
-1	16	4	1	0	0	3	0	0	58	
0	15	0	5	2	0	7	0	1	95	
1	7	0	1	30	4	27	2	0	337	
2	4	1	0	20	42	260	0	0	564	
3	2	0	0	1	10	2760	0	0	771	

Authors express biases even if factuality is underspecified

Aggregate Agreement

Events

Factuality



Agreement improves as we aggregate more non-experts

Models

Event detection - SVM classifier

Features: lemma, POS, event-synset hyponym, brown clusters, dependency paths up to length 1

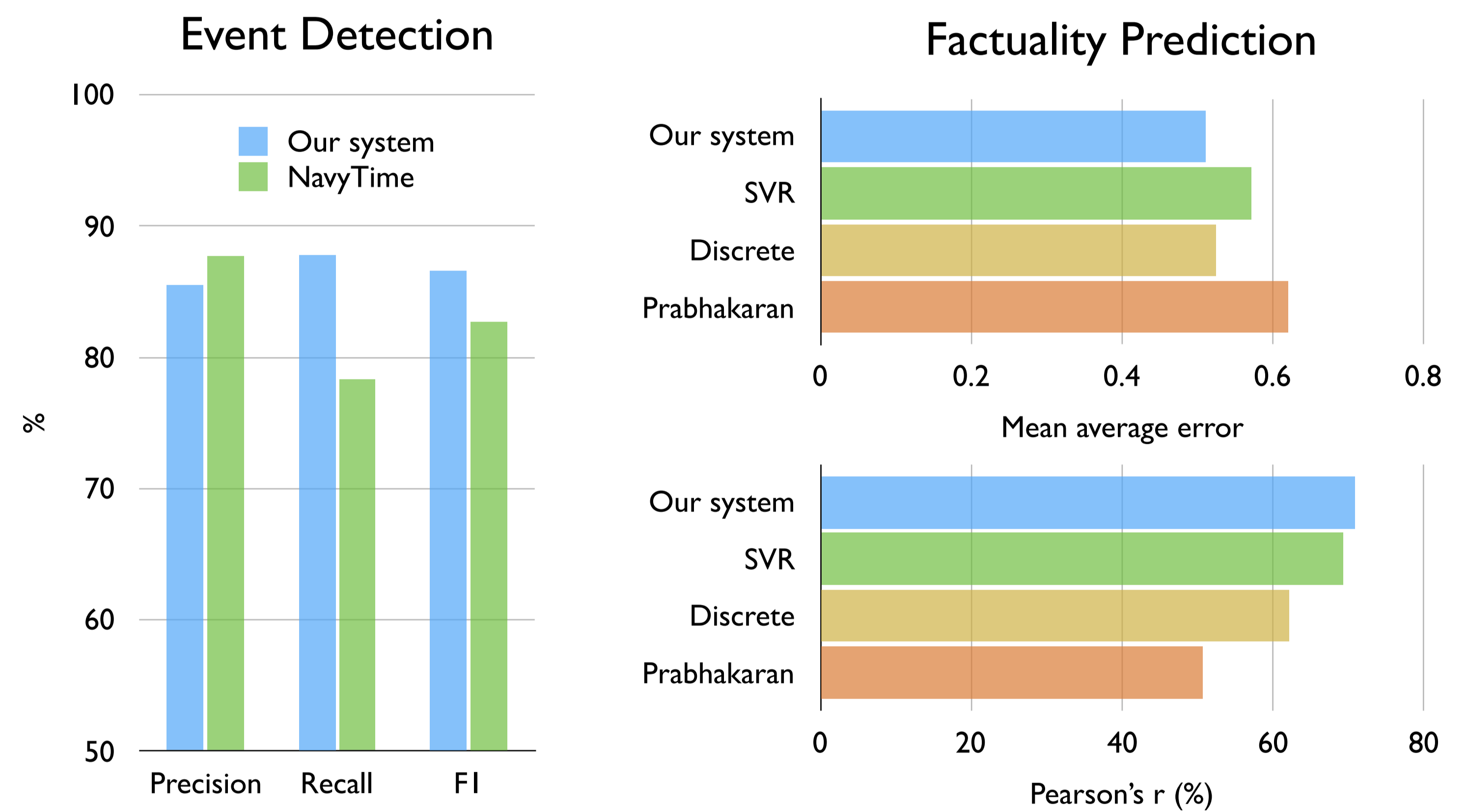
Factuality prediction - Hybrid of SVR and LASSO

$$\min_w \|w\|_1 + C \sum_{i=1}^N \max(0, |y_i - w^T \phi(x_i)| - \epsilon)$$

Features: lemma, POS, dependency paths up to length 2

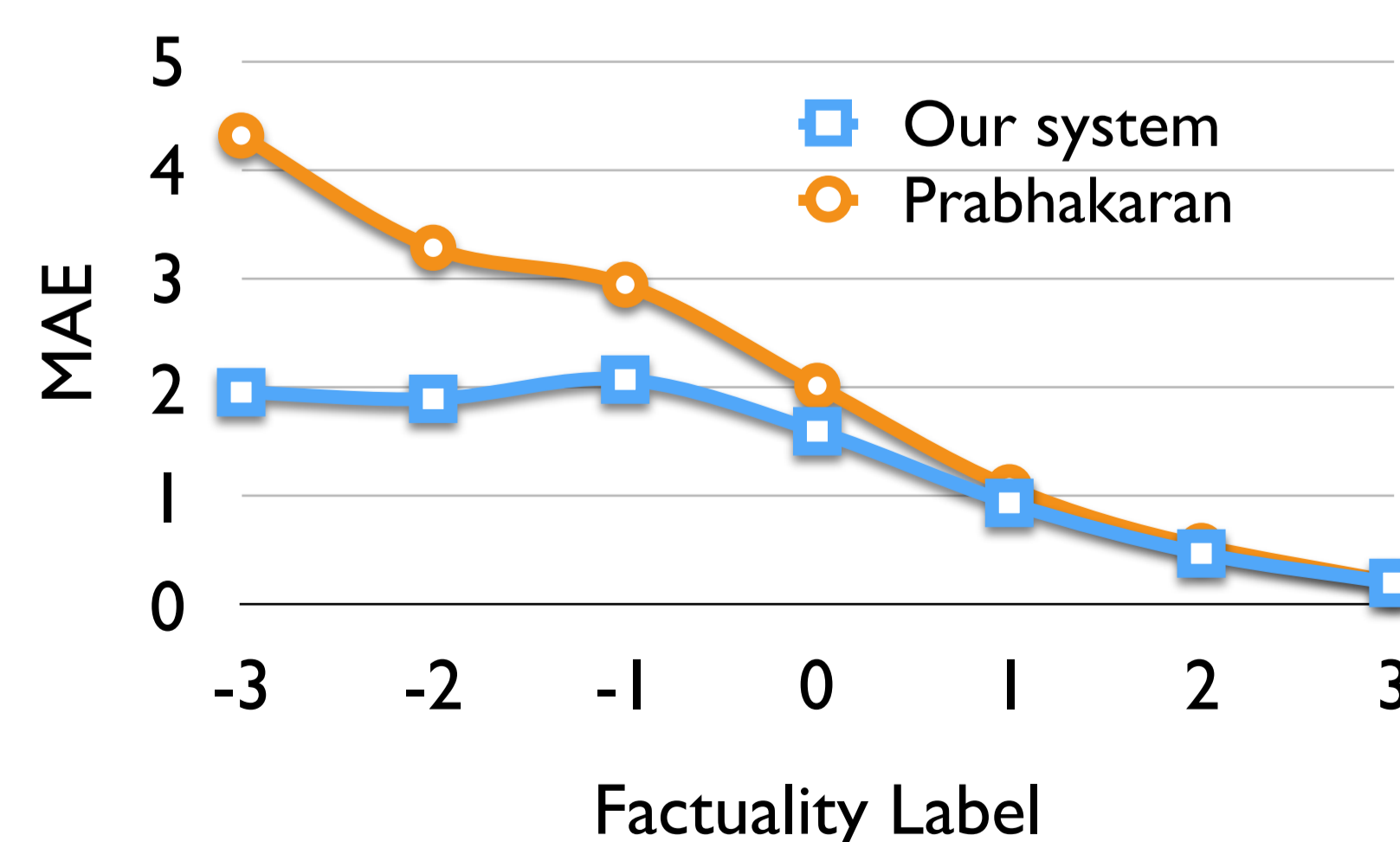
Dependency features capture short-distance event interactions

Results



Error Analysis

Error type	%	Example
Missed lexical cue	64	That <i>sale</i> could still fall through if financing problems develop.
Long distance inference	16	Mesa had <i>rejected</i> a general proposal from StatesWest to combine the two carriers.
World knowledge & pragmatics	12	There was no <i>hint</i> of trouble in the last conversation between controllers and TWA pilot Steven Snyder.



Non-factual events require modeling diverse lexical cues

Related Work

- FactBank: a corpus annotated with event factuality (Sauri and Pustejovsky, 2009)
- Automatic committed belief tagging (Prabhakaran et al., 2010)
- Did it happen? The pragmatic complexity of veridicality assessment (de Marneffe et al. 2012)
- Modeling factuality judgments in social media text (Soni et al. 2014)