Robust Question Answering via Sub-part Alignment
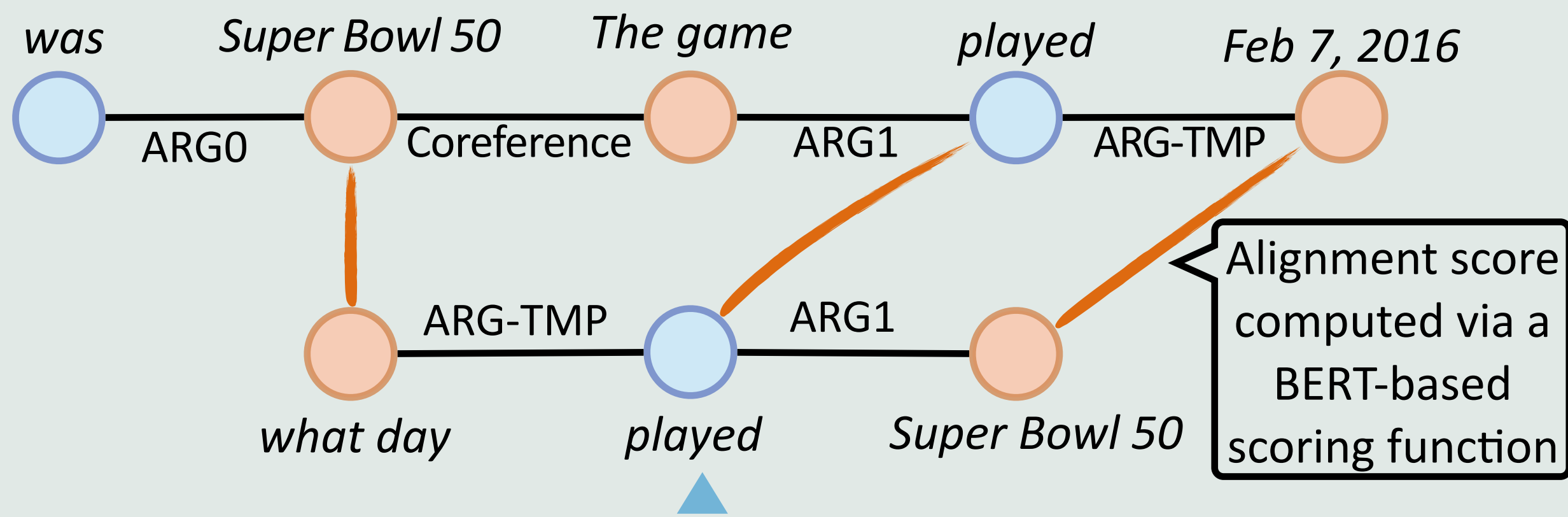
Jifan Chen and Greg Durrett

TAUR LAB @UT Austin

Approach

Make QA explicit by aligning sub-parts

Context: Super Bowl 50 was … The game was played on Feb 7, 2016

| *was* | *Super Bowl 50* | *The game* | *played* | *Feb 7, 2016* |

- *was* —ARG0— *Super Bowl 50* —Coreference— *The game* —ARG1— *played* —ARG-TMP— *Feb 7, 2016*
- *Super Bowl 50* —ARG-TMP— *played* —ARG1— *Super Bowl 50*
- *what day*

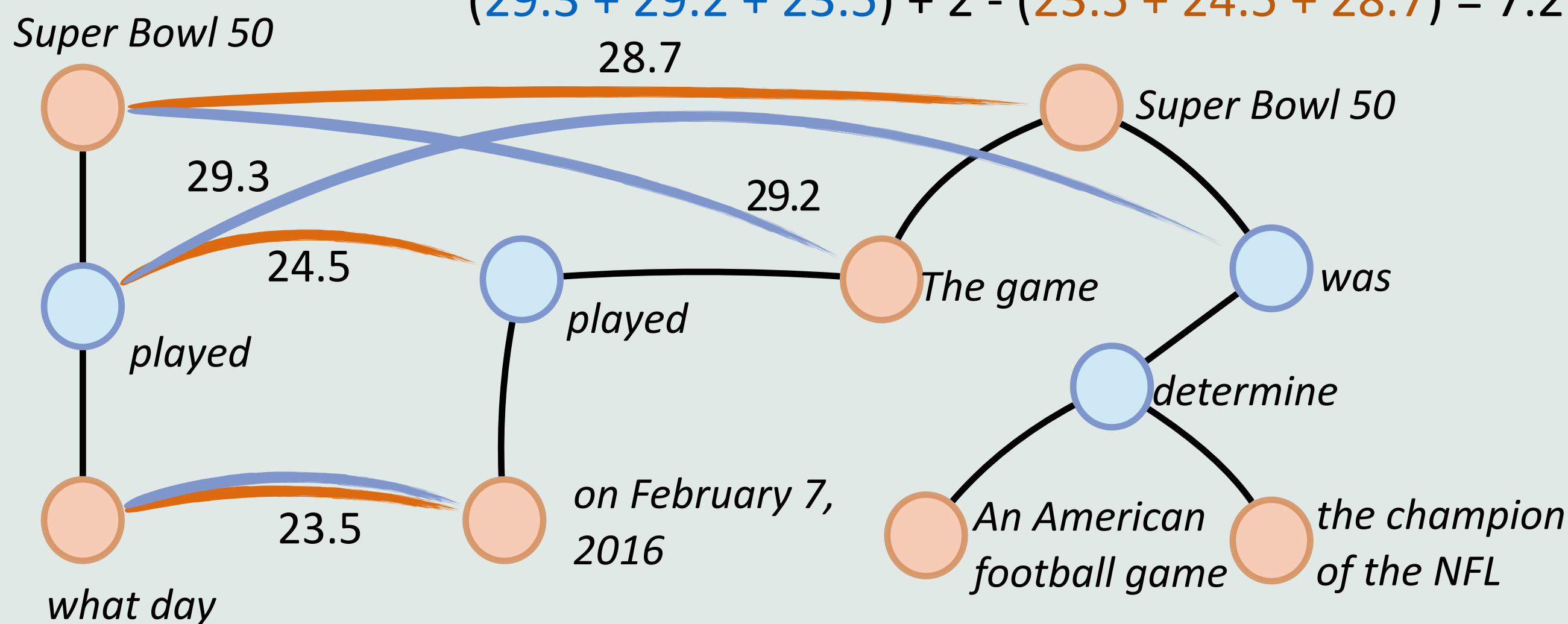**Alignment score computed via a BERT-based scoring function**

Question: What day was Super Bowl 50 played on?
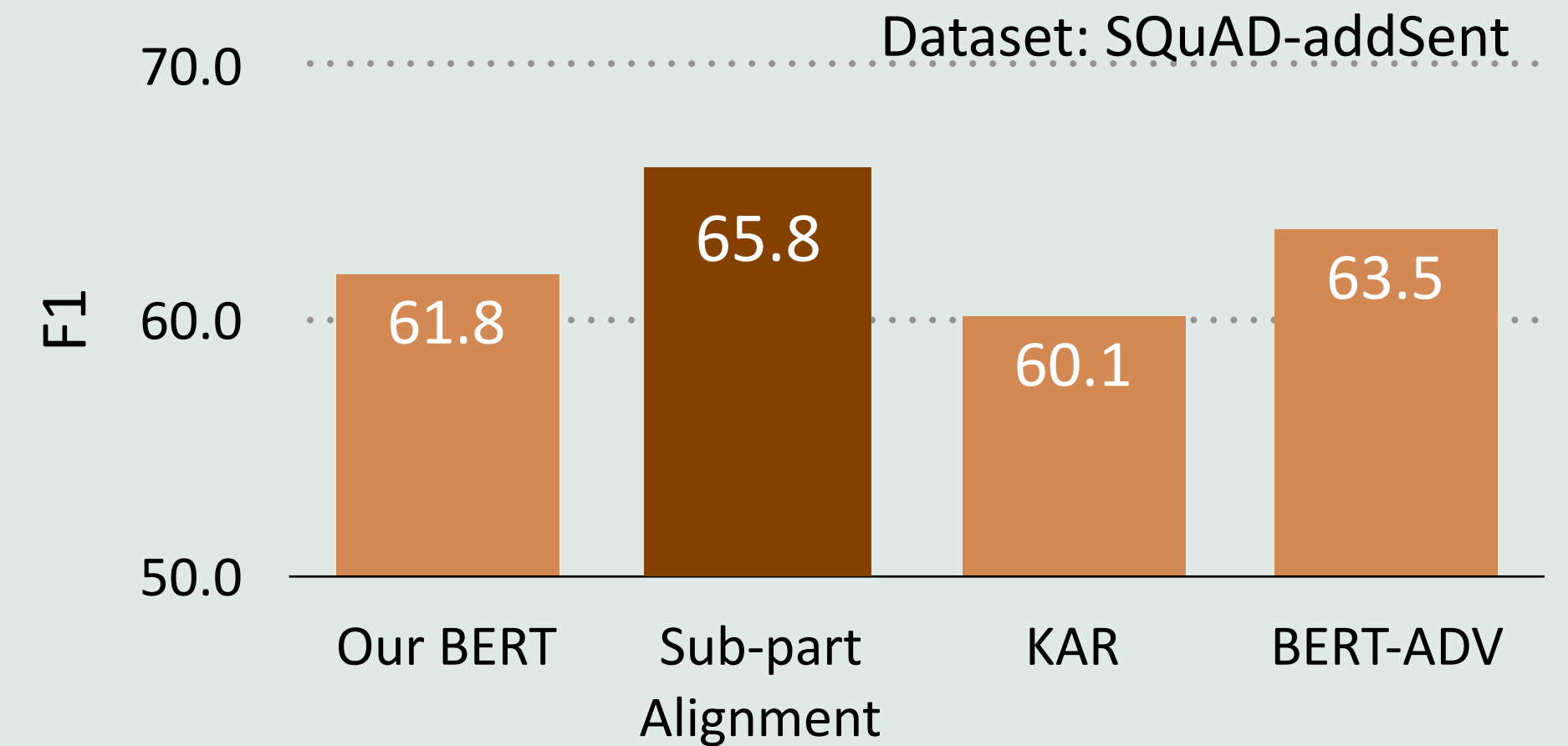
Train as an instance of SSVM

$$L = \max(0, \max(0, \text{predicted score} + \text{Hamming loss}) - \text{gold score})$$

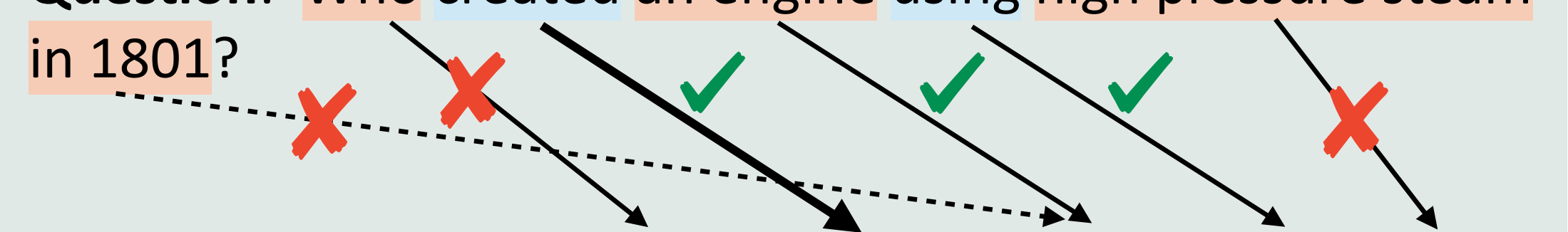$$(29.3 + 29.2 + 23.5) + 2 - (23.5 + 24.5 + 28.7) = 7.2$$

- *Super Bowl 50* — 28.7 — *Super Bowl 50*
- *played* — 29.3 — *The game*
- 29.2 — *was*
- 24.5 — *played*
- *determine*
- 23.5 — *on February 7, 2016*
- *what day*
- *An American football game* — *the champion of the NFL*

What can the model do?

Improve adversarial robustness

Dataset: SQuAD-addSent

	Our BERT	Sub-part Alignment	KAR	BERT-ADV
F1	61.8	65.8	60.1	63.5

Find the unreliable alignment

Question: Who created an engine using high pressure steam in 1801?

✗ ✗ ✓ ✓ ✓ ✗

Adversarial context: Jeff Dean created an engine using low pressure steam in 1790.

Trade coverage for performance

— Sub-part Alignment   - - - BERT

F1 vs Coverage (1–100)