Logical Inference for Counting on Semi-structured Tables

Tomoya Kurosawa and **Hitomi Yanaka**
{kurosawa-tomoya, hyanaka}@is.s.u-tokyo.ac.jp
THE UNIVERSITY OF TOKYO

CONTRIBUTIONS

We propose a **logical inference system** for handling **numerical comparatives** that is based on formal semantics for **NLI between semi-structured tables and texts**.

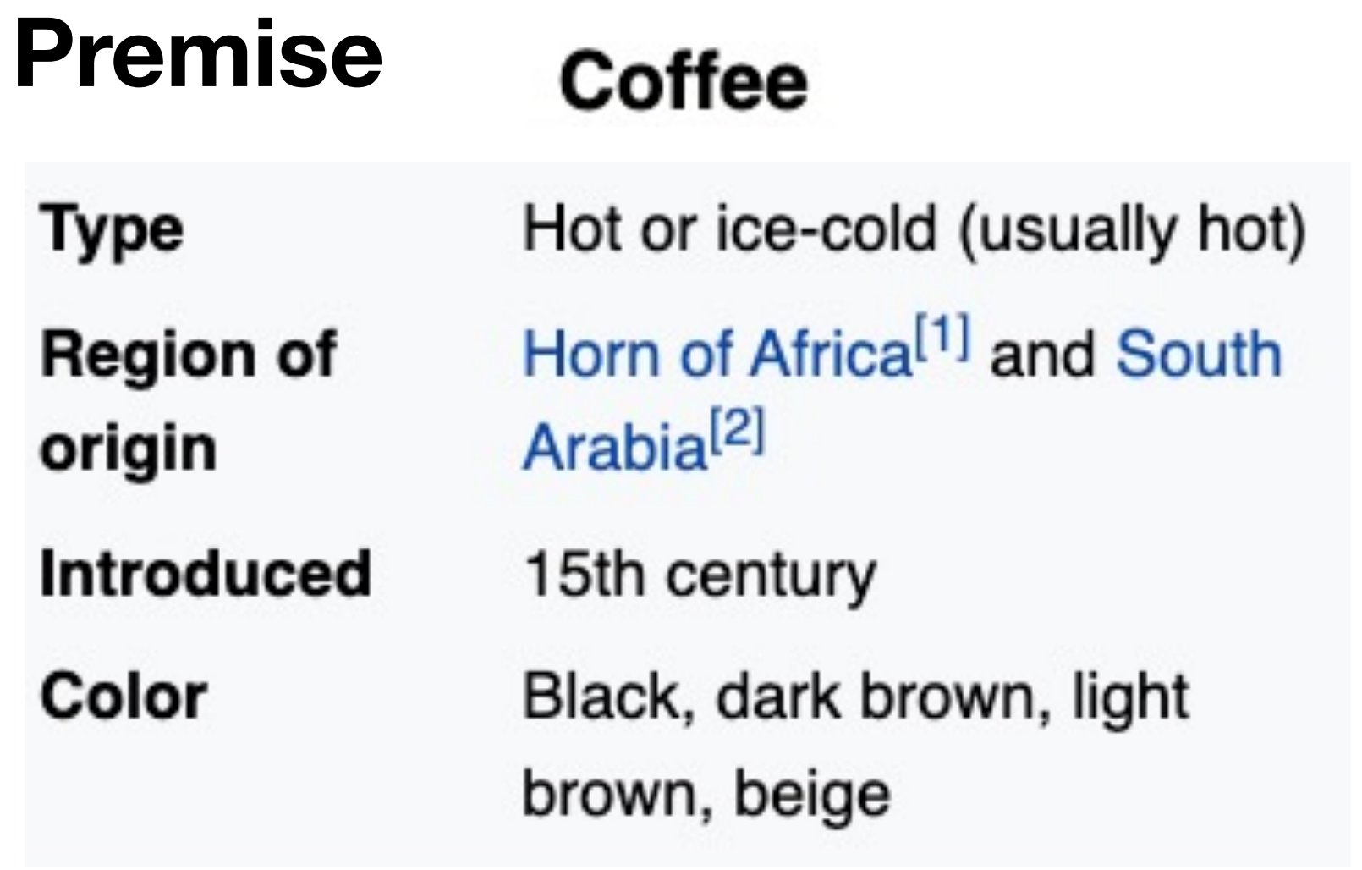
We provide **an evaluation protocol and dataset that focus on numerical comparatives** between semi-structured tables and texts.

We demonstrate **the increased performance of our inference system compared with previous neural network models** on the NLI dataset, **focusing on numerical comparatives** between semi-structured tables and texts.

NLI ON SEMI-STRUCTURED TABLES

The task to determine whether a premise (semi-structured table) entails a hypothesis (sentence) or not.

Premise



Coffee	
Type	Hot or ice-cold (usually hot)
Region of origin	Horn of Africa[1] and South Arabia[2]
Introduced	15th century
Color	Black, dark brown, light brown, beige

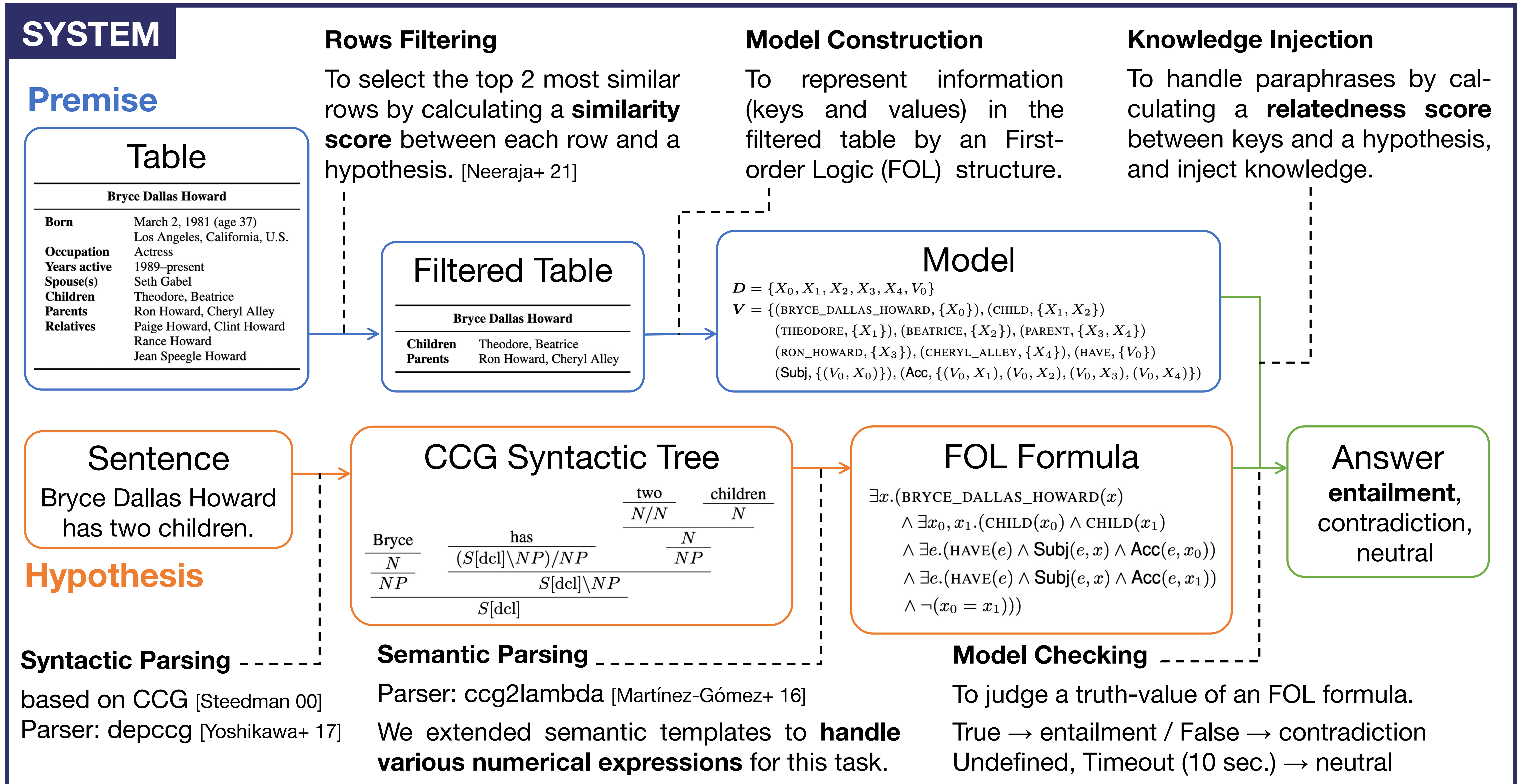
Hypothesis

Coffee has more than four colors.

Answer

Contradiction

SYSTEM



Rows Filtering

To select the top 2 most similar rows by calculating a **similarity score** between each row and a hypothesis. [Neeraja+ 21]

Model Construction

To represent information (keys and values) in the filtered table by an First-order Logic (FOL) structure.

Knowledge Injection

To handle paraphrases by calculating a **relatedness score** between keys and a hypothesis, and inject knowledge.

Premise

Table

Bryce Dallas Howard	
Born	March 2, 1981 (age 37) Los Angeles, California, U.S.
Occupation	Actress
Years active	1989–present
Spouse(s)	Seth Gabel
Children	Theodore, Beatrice
Parents	Ron Howard, Cheryl Alley
Relatives	Paige Howard, Clint Howard Rance Howard Jean Speegle Howard

Filtered Table

Bryce Dallas Howard	
Children	Theodore, Beatrice
Parents	Ron Howard, Cheryl Alley

Model

$$\boldsymbol{D} = \{X_0, X_1, X_2, X_3, X_4, V_0\}$$

$$\begin{aligned}\boldsymbol{V} = \{&(\textsc{bryce\_dallas\_howard}, \{X_0\}), (\textsc{child}, \{X_1, X_2\})\\ &(\textsc{theodore}, \{X_1\}), (\textsc{beatrice}, \{X_2\}), (\textsc{parent}, \{X_3, X_4\})\\ &(\textsc{ron\_howard}, \{X_3\}), (\textsc{cheryl\_alley}, \{X_4\}), (\textsc{have}, \{V_0\})\\ &(\mathsf{Subj}, \{(V_0, X_0)\}), (\mathsf{Acc}, \{(V_0, X_1), (V_0, X_2), (V_0, X_3), (V_0, X_4)\})\end{aligned}$$

Sentence

Bryce Dallas Howard has two children.

Hypothesis

CCG Syntactic Tree

$$\dfrac{\dfrac{\dfrac{\text{Bryce}}{N}}{NP} \quad \dfrac{\dfrac{\text{has}}{(S[\text{dcl}]\backslash NP)/NP} \quad \dfrac{\dfrac{\dfrac{\text{two}}{N/N} \quad \dfrac{\text{children}}{N}}{N}}{NP}}{S[\text{dcl}]\backslash NP}}{S[\text{dcl}]}$$

FOL Formula

$$\begin{aligned}&\exists x.(\textsc{bryce\_dallas\_howard}(x)\\ &\quad \wedge \exists x_0, x_1.(\textsc{child}(x_0) \wedge \textsc{child}(x_1)\\ &\quad \wedge \exists e.(\textsc{have}(e) \wedge \mathsf{Subj}(e, x) \wedge \mathsf{Acc}(e, x_0))\\ &\quad \wedge \exists e.(\textsc{have}(e) \wedge \mathsf{Subj}(e, x) \wedge \mathsf{Acc}(e, x_1))\\ &\quad \wedge \neg(x_0 = x_1)))\end{aligned}$$

Answer

entailment, contradiction, neutral

Syntactic Parsing

based on CCG [Steedman 00]
Parser: depccg [Yoshikawa+ 17]

Semantic Parsing

Parser: ccg2lambda [Martínez-Gómez+ 16]

We extended semantic templates to **handle various numerical expressions** for this task.

Model Checking

To judge a truth-value of an FOL formula.

True → entailment / False → contradiction
Undefined, Timeout (10 sec.) → neutral

DATASET CREATION

We created a new dataset (105 problem sets; 1,979 test cases) for the numerical understanding of semi-structured tables by extracting from InfoTabS [Gupta+ 20] because

- the number of test cases for numerical understanding is limited to InfoTabS
- to evaluate whether NLI systems consistently perform inference with numerical comparatives involving various numbers

① A part of **Premise**

Johnny Depp	
Spouse(s)	Lori Allison, Amber Heard

② **Base Hypothesis**

Johnny Depp has been married **exactly twice**.
(actual number n = 2)

③ **Testcases**

Johnny Depp has been married
+
Numerical comparatives
ϕ, less than, no more than, exactly,
at least, no less than, more than
+
Numerals for the number of entities / events
once (n – 1), twice (n), three times (n + 1)

RESULT

We compare our system with +KG explicit [Neeraja+ 21], previous neural network-based approach.

+KG explicit makes sentence representations of tables and uses **RoBERTa-large** [Liu+ 19] for encoding premise-hypothesis pairs.

Using our dataset, we observed that **our system performed more robustly** than the previous neural network-based model.

The accuracy of problem sets whose test cases were all predicted correctly.

	+KG	Ours
All problem sets	0.03	**0.31**
entailment & contradiction	0.00	**0.27**

The accuracy for each numerical comparative construction. k indicates a number.

	+KG	Ours
less than k	0.10	**0.36**
no more than k	0.10	**0.35**
exactly k	0.19	**0.32**
k	0.24	**0.33**
at least k	0.08	**0.32**
no less than k	0.19	**0.33**
more than k	0.17	**0.35**

Average and maximum run time (sec.) for model checking with and without optimization.

Optim.	Avg.	Max.
-	3.20	185.17
+	**0.04**	**1.26**

We **optimize the NLTK program** for model checking to **make judgments faster** by

- sorting variables
- avoiding some substitution

LINKS

ynklab/sst_count @m_m_ast_