TRANS-KBLSTM: An External Knowledge Enhanced Transformer BiLSTM model for Tabular Reasoning

https://trans-kblstm.github.io

Yerram Varun[1*], Aayush Sharma[1*], Vivek Gupta[2*],
[1]IIT Guwahati; [2]University of Utah

[2]Bloomberg Ph.D. Fellow
[2]on academic job market

TABULAR INFERENCE PROBLEM

- The **tabular natural language inference** problem is similar to standard NLI
- But here, the **premises are tabular data**
- Task: to decide whether given hypothesis is **true** (entailment), **false** (contradiction) or **undetermined** (neutral) given a premise table

New York Stock Exchange	
Type	Stock exchange
Location	New York City, New York, U.S.
Founded	May 17, 1792; 226 years ago
Currency	United States dollar
No. of listings	2,400
Volume	US$20.161 trillion (2011)

H1: NYSE has fewer than 3,000 stocks listed.
H2: Over 2,500 stocks are listed in the NYSE.
H3: S&P 500 stock trading volume is over $10 trillion.

In this example from the InfoTabS dataset (Gupta et al., 2020),

H1: entailed ; **H2:** contradictory ; **H3:** neutral

Check out InfoTabs (Gupta et al., 2020)
https://infotabs.github.io

KNOWLEDGE ADDITION

- Many a times, **External knowledge** is necessary for model inference.

- These requirements limit the performance of neural models

- **Task:** To use external **knowledge graphs** to supplement deep learning architectures for improved reasoning.

In this example from INFOTABS, predicting the Gold label requires broad understanding of

California is located on the Coast.

	James Hetfield
Birth Name	James Alan Hetfield
Born	Aug. 3, 1963(age 58), California, U.S.
Genres	Heavy metal, thrash metal, hard rock
Occupation(s)	Musician, Singer
Instruments	Vocals, Guitar
Years active	1978-present
Labels	Warner Bros, Elektra, MegaForce
Hypothesis	James Hetfield was born on the west coast of the USA.
Focused Relation	coast $\xleftarrow{AtLocation}$ california
Human	Entailment
RoBERTa	Neutral
Trans-KBLSTM	Entailment

MOTIVATION

Recent work on using external knowledge for tabular reasoning use **explicit addition of knowledge** i.e. knowledge appended at additional input context.

MOTIVATION

Recent work on using external knowledge for tabular reasoning use **explicit addition of knowledge** i.e. knowledge appended at additional input context.

Questions

- ***Knowledge Extraction*: ** How can we extract **contextually relevant** knowledge from external source?

MOTIVATION

Recent work on using external knowledge for tabular reasoning use **explicit addition of knowledge** i.e. knowledge appended at additional input context.

Questions

- ***Knowledge Extraction*:** How can we extract **contextually relevant** knowledge from external source?
- ***Knowledge Representation*:** How to **effectively represent** external **semantic knowledge relations**?

MOTIVATION

Recent work on using external knowledge for tabular reasoning use **explicit addition of knowledge** i.e. knowledge appended at additional input context.

Questions

- ***Knowledge Extraction*****:** How can we extract **contextually relevant** knowledge from external source?
- ***Knowledge Representation*****:** How to **effectively represent** external **semantic knowledge relations**?
- ***Knowledge Integration*****:** How to **schematically integrate external knowledge** into model architectures?

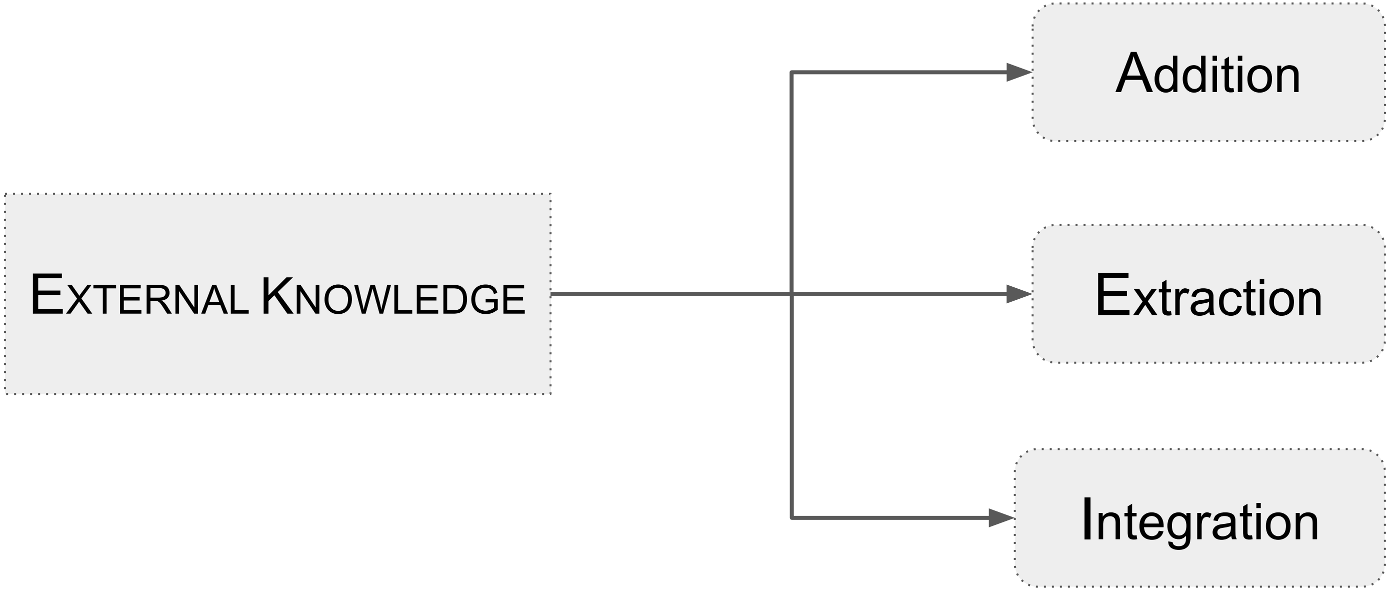
TAKEAWAY

Through a ***novel architecture, Trans-KBLSTM***, this work investigates strategies to tackle challenges inherent in **existing methodologies** of ***Knowledge Extraction, Addition, and Integration.***

The effectiveness is assessed through **INFOTABS**, a Tabular NLI Dataset.

Check out InfoTabs (Gupta et al., 2020) https://infotabs.github.io

CHALLENGES



Challenges: Knowledge Extraction

- KG-Explicit (Neeraja et al., 2021) augments the input with **lengthy key definitions**.

- Add **noise** and **confusion** caused by **lengthy additions**. At times definitions are incorrect.

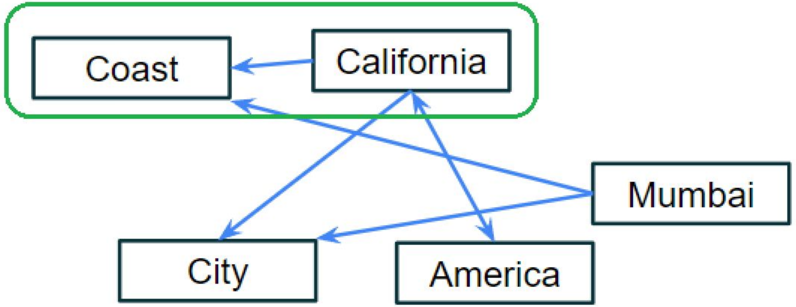
Dr. Max Born has no connection with Julius Caesar

Orignal Premise Julius Caesar was born on 12 or 13 July 100 BC Rome. Julius Caesar died on 15 March 44 BC (aged 55) Rome. The resting place of Julius Caesar is Temple of Caesar, Rome. The spouse(s) of Julius Caesar are Cornelia (84-69 BC; her death), Pompeia (67-61 BC; divorced), Calpurnia (59-44 BC; his death).

Orignal Premise + KG explicit Julius Caesar died on 15 March 44 BC (aged 55) Rome. **The resting place of Julius Caesar is Temple of Caesar, Rome.** Julius Caesar was born on 12 or 13 July 100 BC Rome. The spouse(s) of Julius Caesar are Cornelia (84-69 BC; her death), Pompeia (67-61 BC; divorced), Calpurnia (59-44 BC; his death). KEY: Died is defined as pass from physical life and lose all bodily attributes and functions necessary to sustain life . KEY: Resting place is defined as a cemetery or graveyard is a place where the remains of dead people are buried or otherwise interred . KEY: Born is defined as british nuclear physicist (born in germany) honored for his contributions to quantum mechanics (1882-1970) . KEY: Spouse is defined as a spouse is a significant other in a marriage, civil union, or common-law marriage .

Hypothesis Julius Caesar was buried in Rome.

SOLUTION: RELATIONAL CONNECTIONS AND KGs



sample knowledge graph

James Hetfield	
Birth Name	James Alan Hetfield
Born	Aug. 3, 1963(age 58), California, U.S.
Genres	Heavy metal, thrash metal, hard rock
Occupation(s)	Musician, Singer
Instruments	Vocals, Guitar
Years active	1978-present
Labels	Warner Bros, Elektra, MegaForce
Hypothesis	James Hetfield was born on the west coast of the USA.

table premise relevant attention

Semantic knowledge graphs represent the relationships between the hypothesis and premise token pairs.

To extract relevant knowledge, use the semantic relational connections between premise and hypothesis tokens.

CHALLENGES: KNOWLEDGE ADDITION

Definition adds lengthy text to the multi-head attention.

Unnecessary noise is introduced in this process

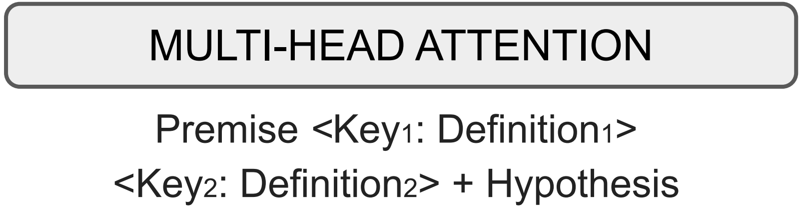
MULTI-HEAD ATTENTION

Premise <Key$_1$: Definition$_1$> <Key$_2$: Definition$_2$> + Hypothesis

Challenges: Knowledge Addition

Definition adds lengthy text to the multi-head attention.

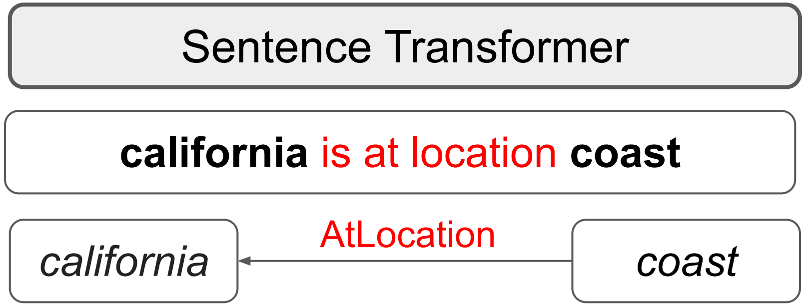
Unnecessary noise is introduced in this process



Solution

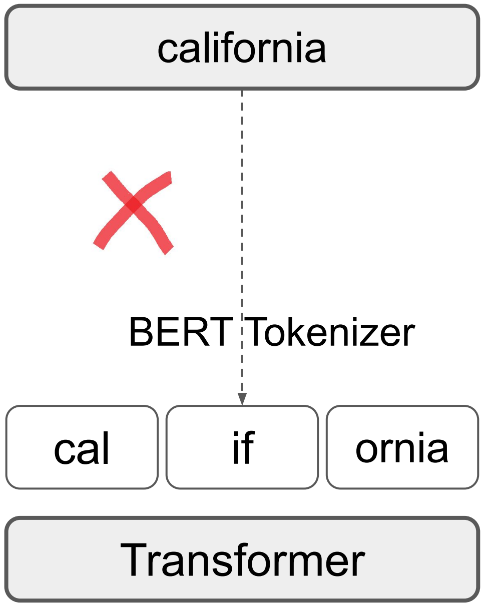
Knowledge Triplets are converted to sentences.

Sentences are encoded using Sentence Transformers.

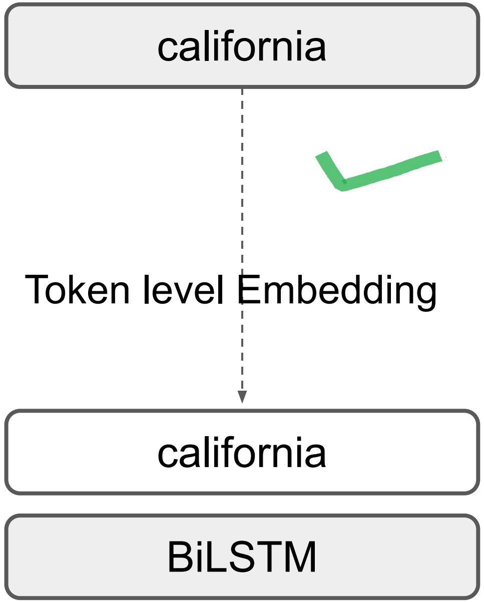


CHALLENGES : KNOWLEDGE INTEGRATION

Word Pair External Knowledge Relations are *not compatible* with ***tokenized transformer inputs.***



SOLUTION: USING BiLSTM MODELS



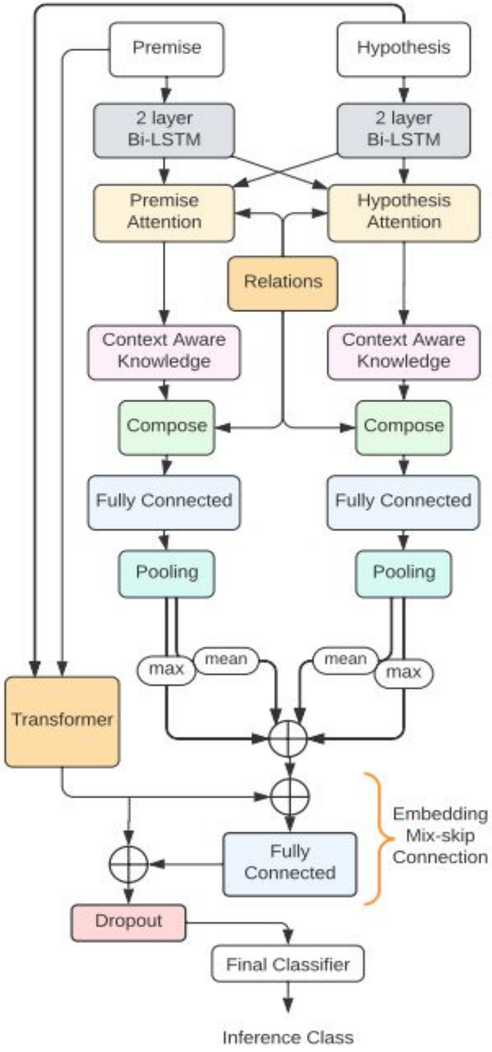
Use BiLSTM Models !!

BiLSTMs employ token level embeddings, thus **complete word pair relations**.

PROPOSED APPROACH

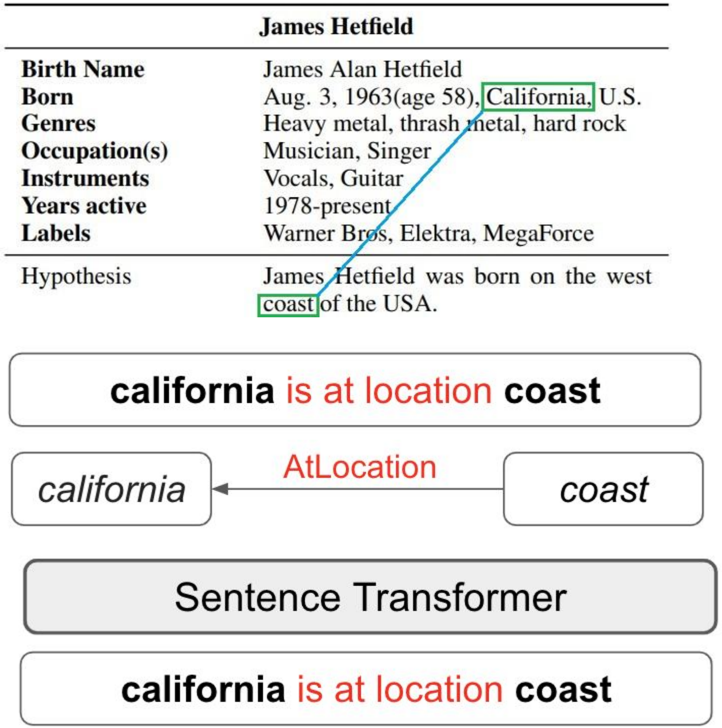
TRANS-KBLSTM

An Overview of the Architecture



PREPROCESSING

- Retrieve relational connections
- Convert into sentence triplets
- Encode using Sentence transformers



James Hetfield	
Birth Name	James Alan Hetfield
Born	Aug. 3, 1963(age 58), California, U.S.
Genres	Heavy metal, thrash metal, hard rock
Occupation(s)	Musician, Singer
Instruments	Vocals, Guitar
Years active	1978-present
Labels	Warner Bros, Elektra, MegaForce
Hypothesis	James Hetfield was born on the west coast of the USA.

Relations Attention And Embedding

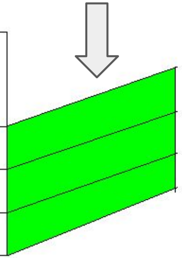
Premise Hypothesis	a1 I	a2 Like	a3 You
b1 I	r11	r21	r31
b2 Hate	r12	r22	r32
b3 You	r13	r23	r33

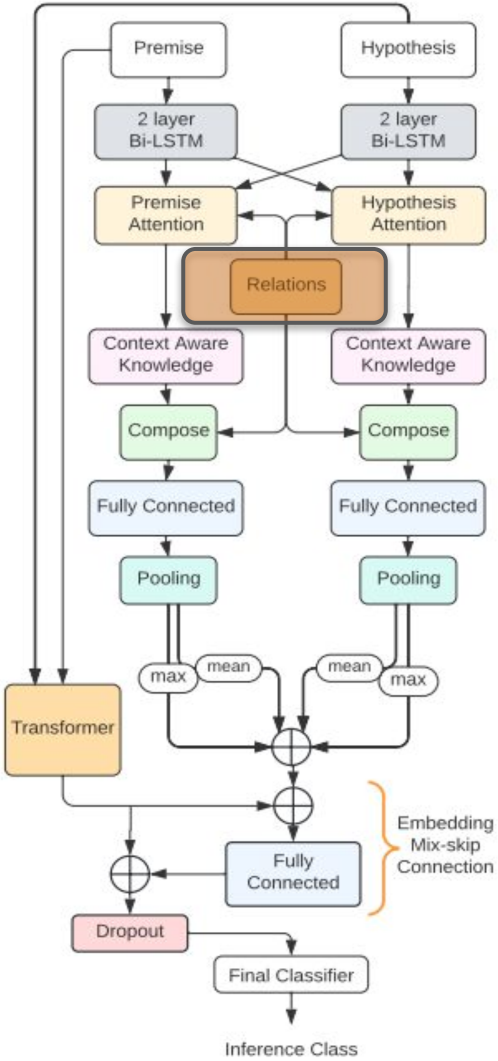
Relational Attention Matrix

Relational Embedding Matrix

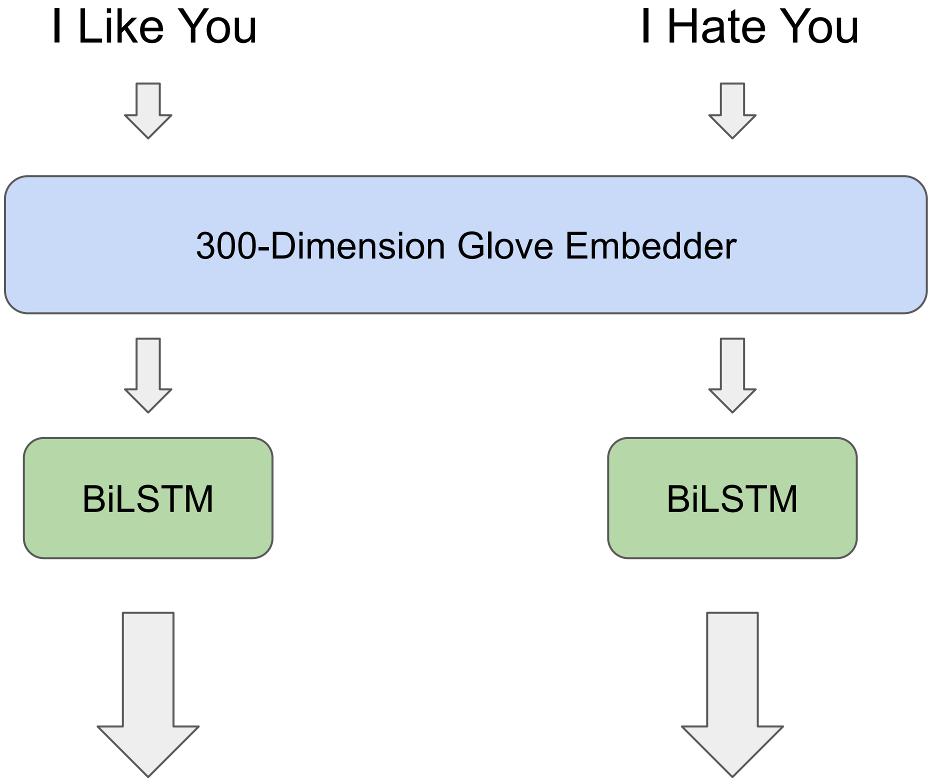
BERT Representations

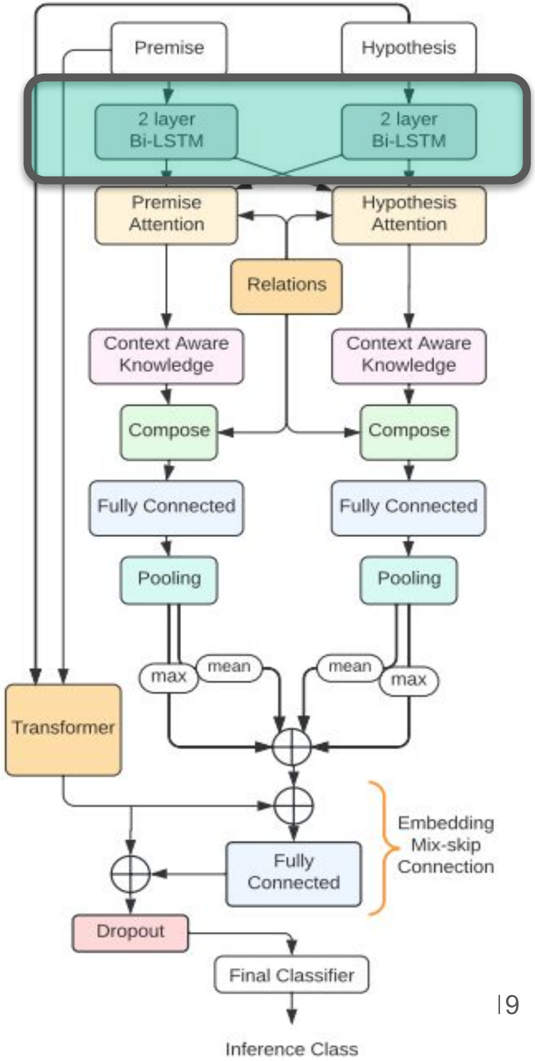
Premise Hypothesis	a1 I	a2 Like	a3 You
b1 I	R11	R21	R31
b2 Hate	R12	R22	R32
b3 You	R13	R23	R33



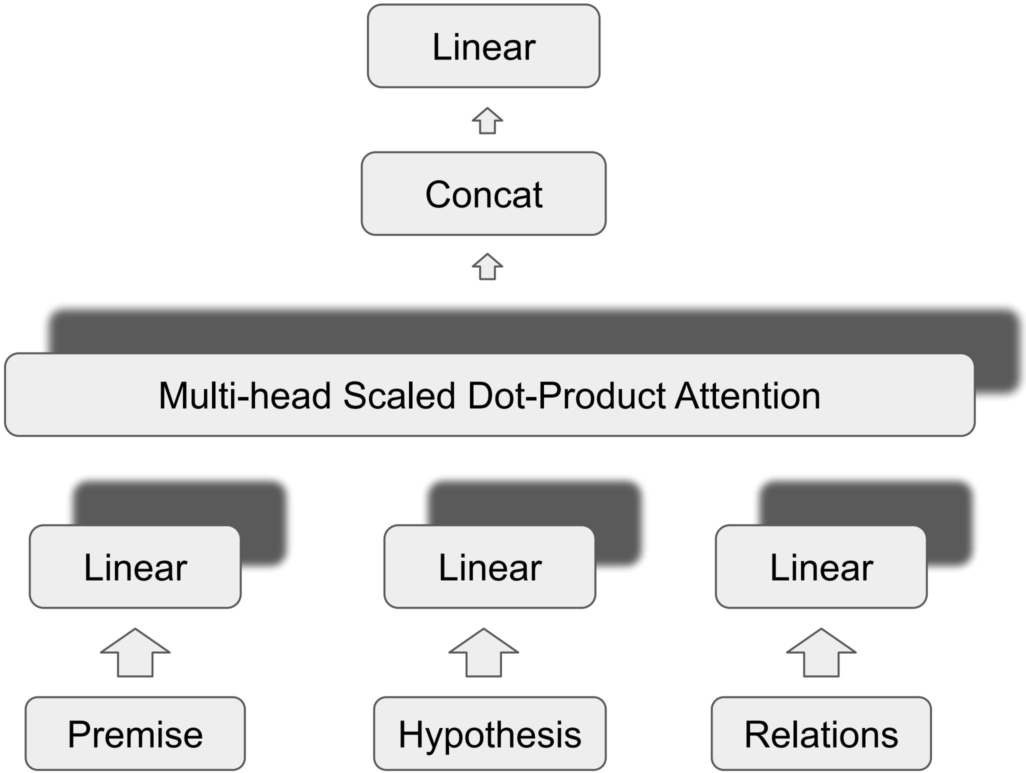


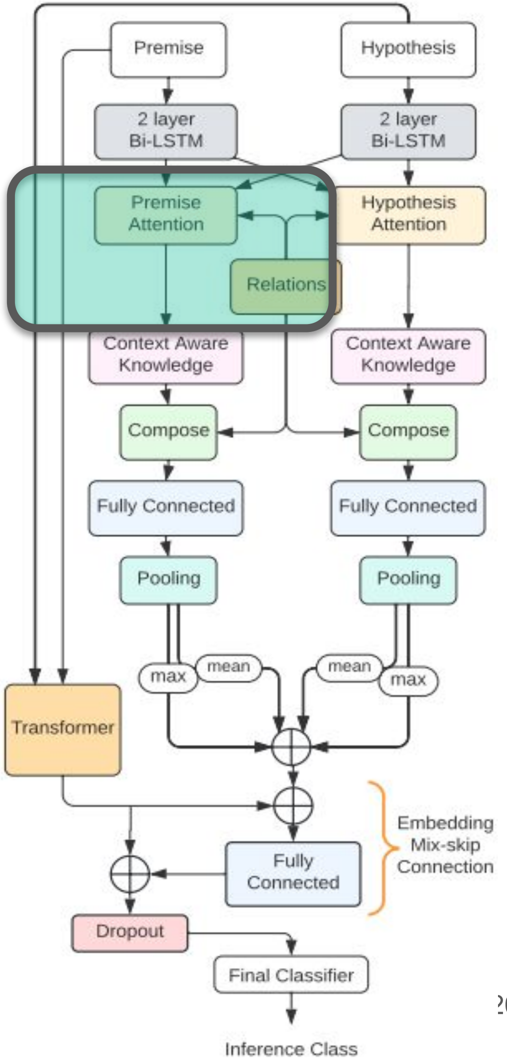
BiLSTM Encoders



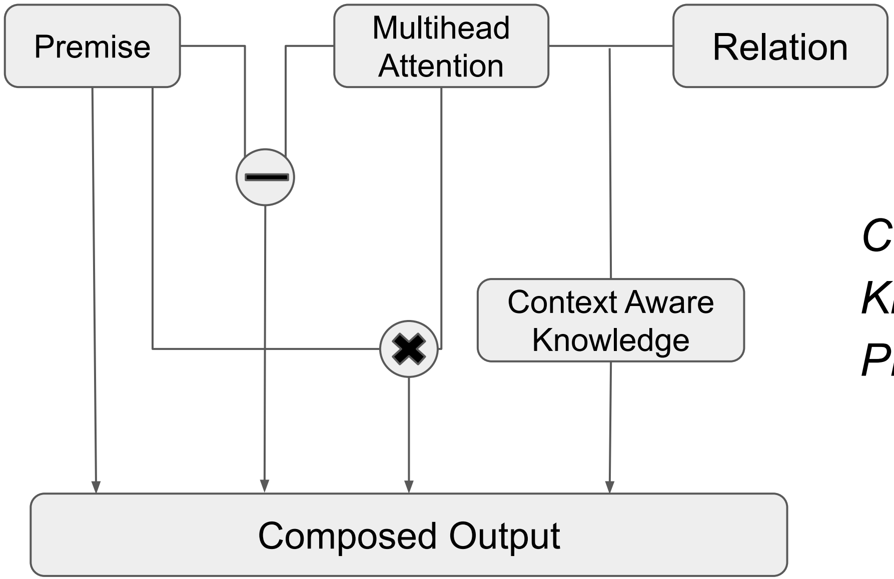


PROPOSED APPROACH: ATTENTION HEADS

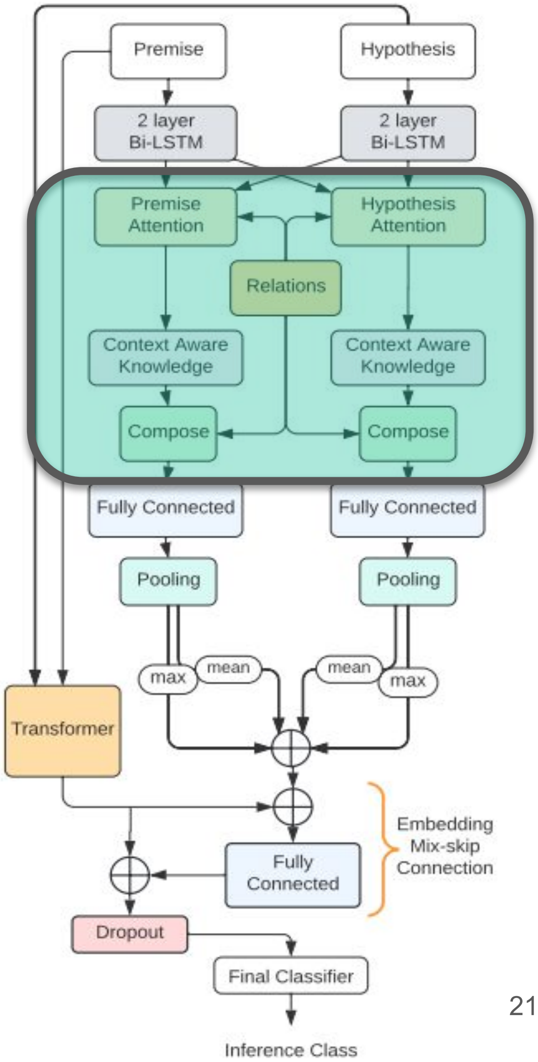




COMPOSE KNOWLEDGE



Composition of *Knowledge* with *Pr/Hyp Attention*

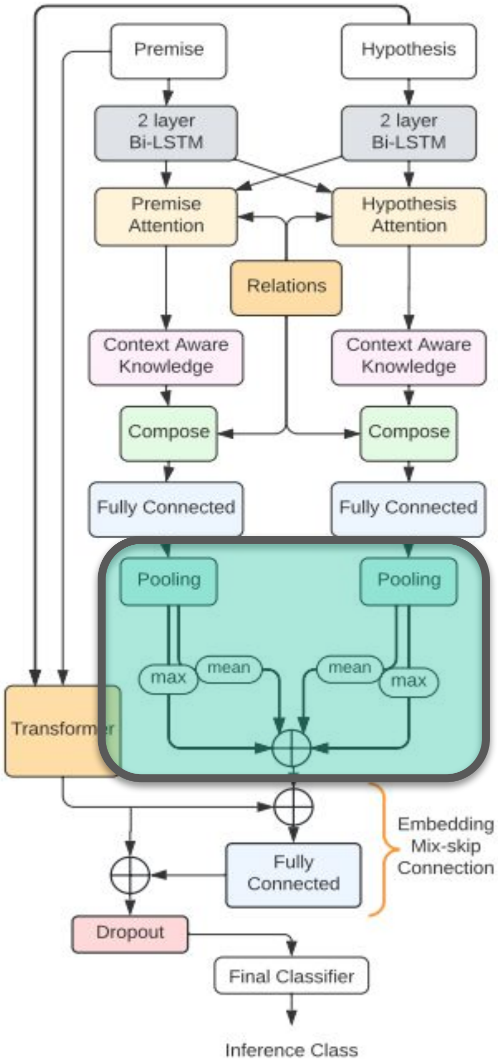


MEAN AND MAX POOLING

The Composed Premise and Hypothesis are

MEAN and **MAX** pooled separately.

$$p_{mean} = \mathrm{MeanPool}(p^m)\ ;\ p_{max} = \mathrm{MaxPool}(p^m)$$
$$h_{mean} = \mathrm{MeanPool}(h^m)\ ;\ h_{max} = \mathrm{MaxPool}(h^m)$$

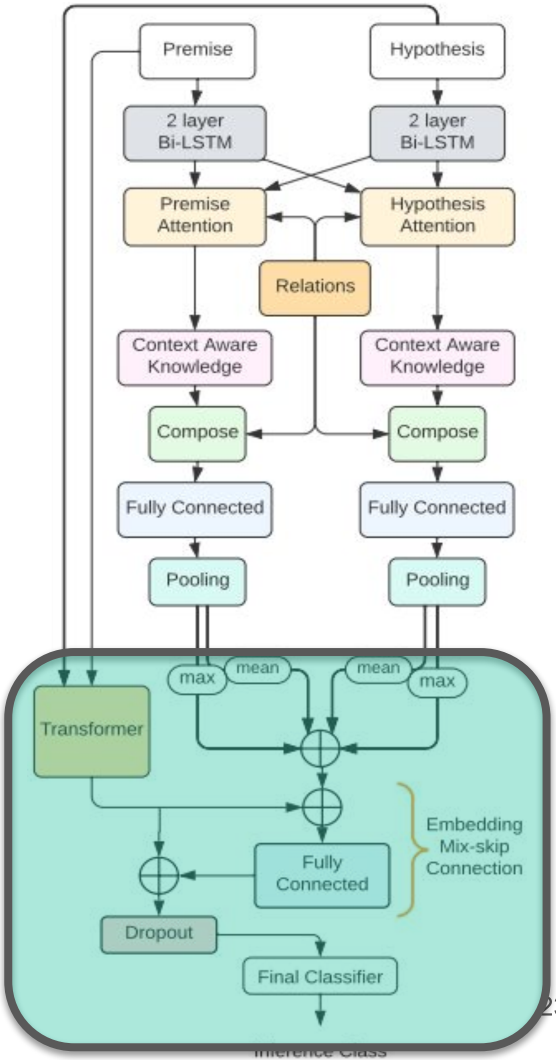


TRANSFORMER

Embedding Mix-Skip Connection

- Mix representations from transformers and the pooling Layer.
- A effective way to integrate both the embeddings.

- Apply Dropout for Regularization
- Fully connected layers
- Final Dense layer with 3 class outputs.



INFOTABS DATASET

InfoTabS dataset splits :

- **α1** contains table from same domain (similar to dev & train set)
- **α2** has examples from same domain but entail-contradict label (e.g. ‘over’ to ‘under’) flipped by minimal change i.e. **adversarial**.
- **α3** is **zero-shot** cross domain tables (exclusive from train set domains)

Check out InfoTabs: https://infotabs.github.io

Models

We considered the following models for our experiments:

- **RoBERTa** : a baseline Transformer model without knowledge.
- **KG Explicit** represent Knowledge-InfoTabS (Neeraja et al., 2021)
- **Tok-KTrans** appending *WordNet Tokens* to Transformer Inputs
- **Trans-KBLSTM** our new proposed approach

Check out Knowledge-InfoTabs: https://knowledge-infotabs.github.io

RESULTS AND ANALYSIS

- How effective is our proposed approach for INFOTABS for:
  - *Full Supervision*
  - *Limited Supervision*
- Is our proposed approach *more beneficial* to certain sorts of reasoning types?
- Ablation Study: How important are each individual components?
  - *Embedding mix-skip Connection*
  - *Knowledge Addition*
  - *Independent training of Transformer and LSTM*
  - MNLI Pretraining
  - Transformer Size

refer to paper for details

RESULTS AND ANALYSIS

- How effective is our proposed approach for INFOTABS for:
  - *Full Supervision*

Model	Dev	$\alpha1$	$\alpha2$	$\alpha3$
w/o Knowledge	77.30	76.44	70.49	69.05
Tok-KTrans	78.17	76.19	70.75	69.77
KG Explicit	78.97	77.84	71.13	69.58
Trans-KBLSTM	**79.92**	**79.62**	**72.10**	**70.21**

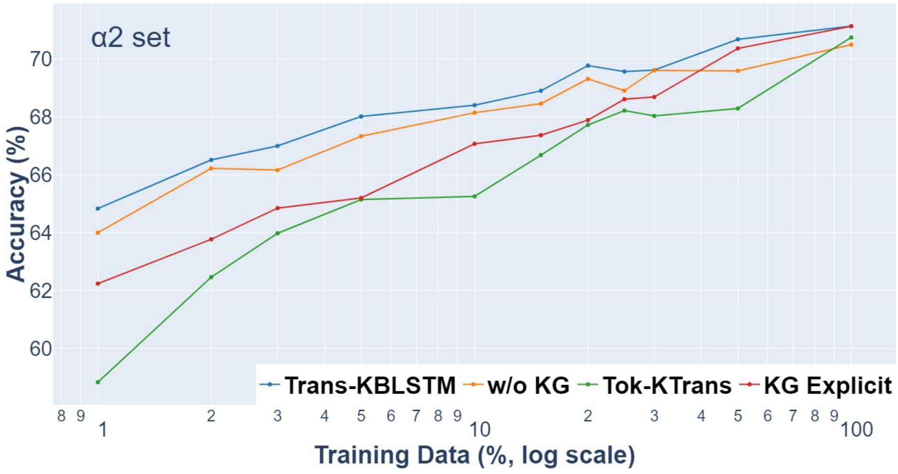
Our proposed approach outperform other baselines!

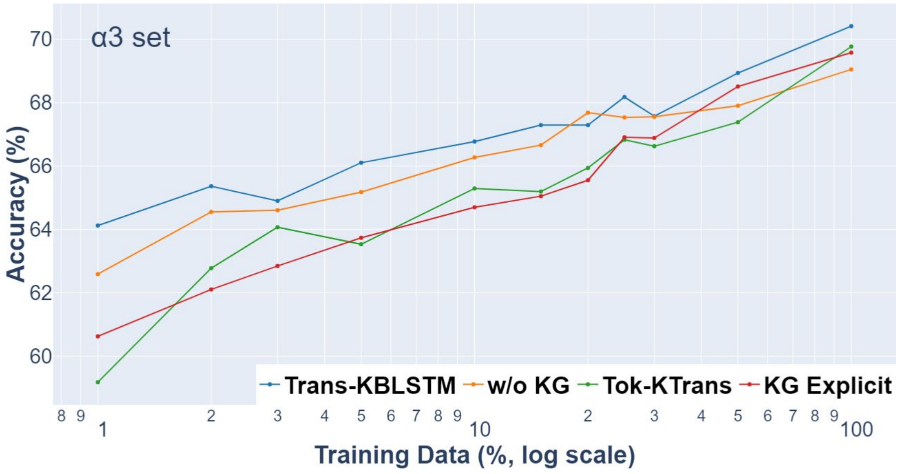
**Reported numbers are average over three random seed runs*

Results And Analysis

- How effective is our proposed approach for InfoTabS for:
  - *Full Supervision*
  - *Limited Supervision*

Limited Supervision Setting

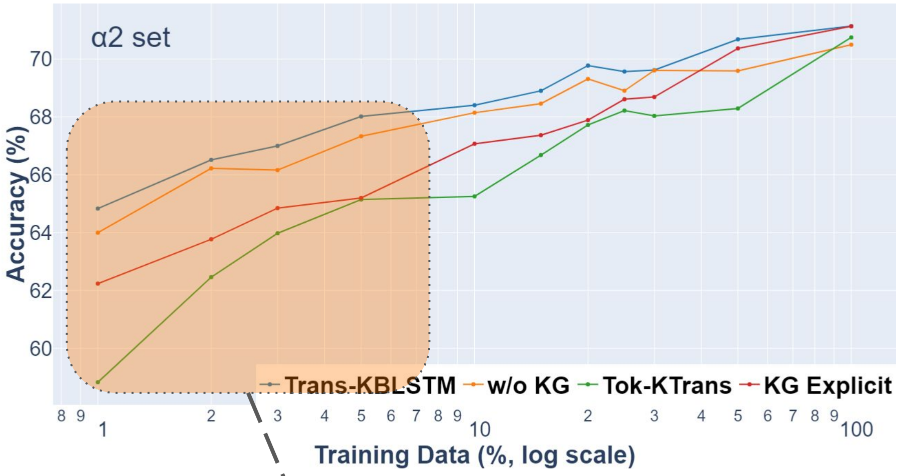


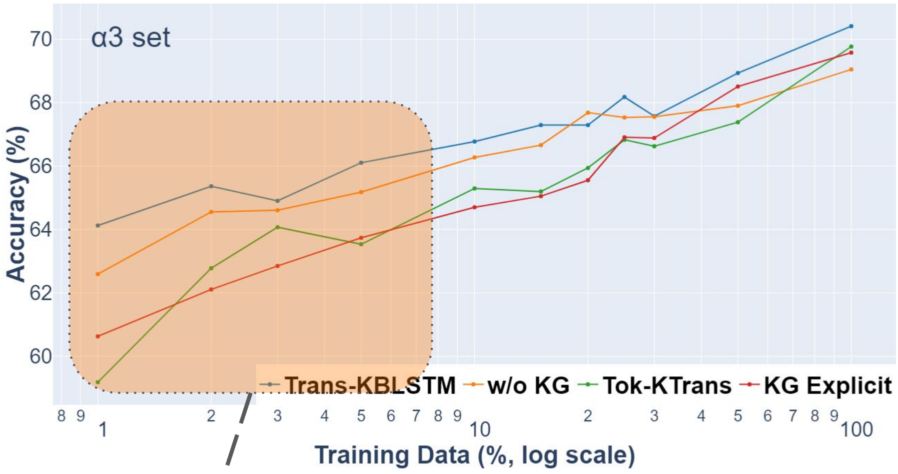


Model performance trained with limited supervision for α_2 and α_3

**Reported numbers are average over three random seed runs*

LIMITED SUPERVISION SETTING



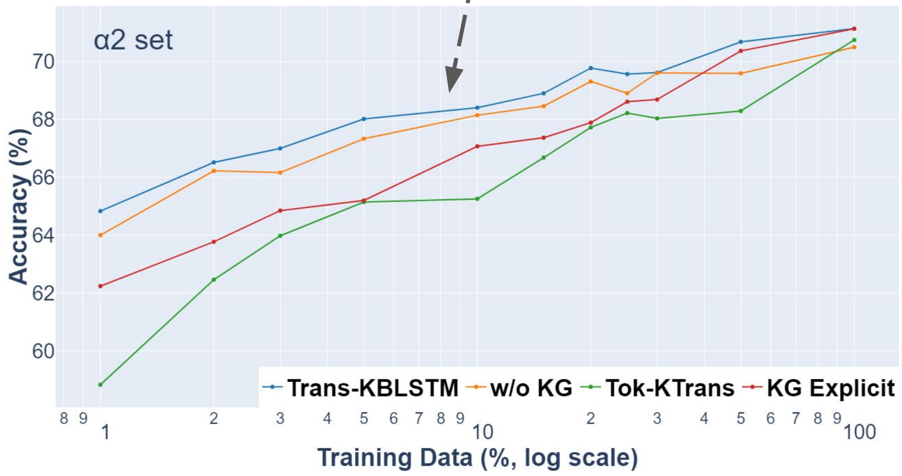


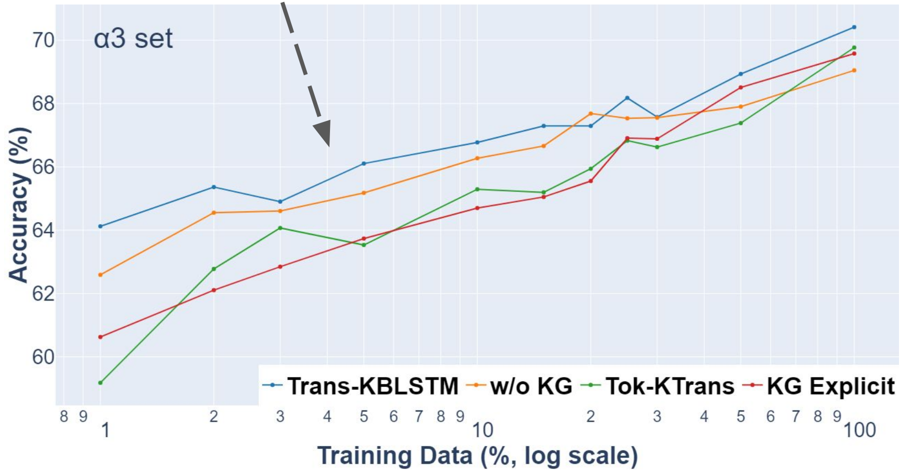
In comparison to complete supervision, improvement in limited setting more substantial.

**Reported numbers are average over three random seed runs*

RESULTS AND ANALYSIS: LIMITED SUPERVISION SETTING

Trans-KBLSTM outperform other baseline models!



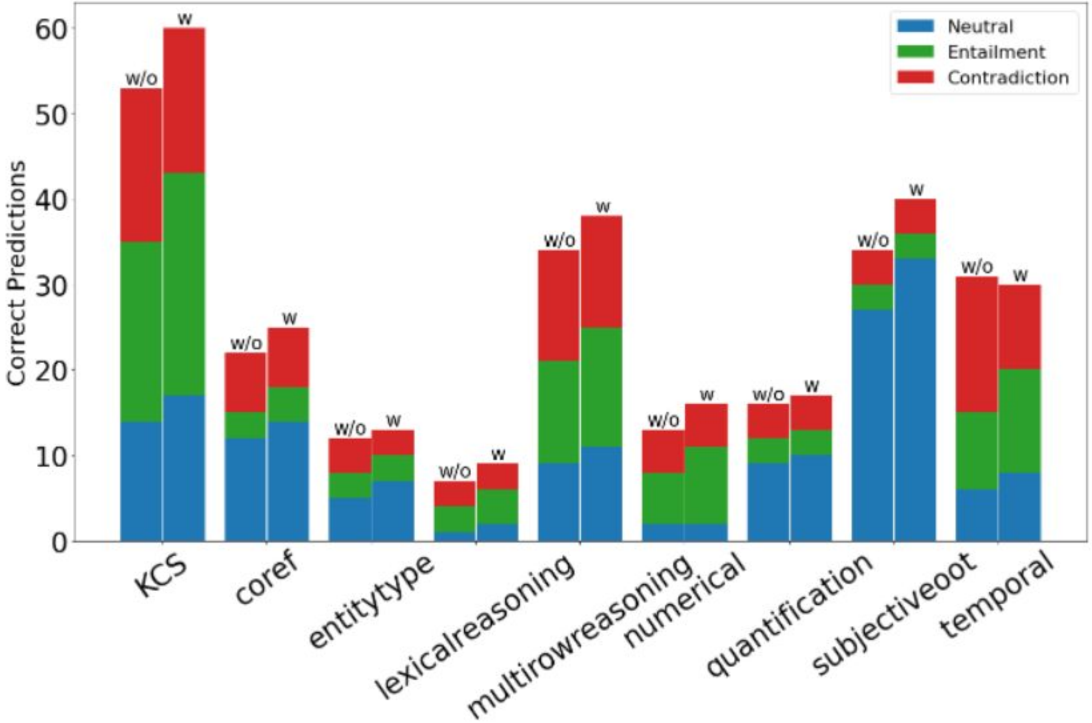


**Reported numbers are average over three random seed runs*

RESULTS AND ANALYSIS

- How effective is our proposed approach for INFOTABS for:
  - *Full Supervision*
  - *Limited Supervision*
- Is our proposed approach *more beneficial* to certain sorts of reasoning types?

Reasoning Analysis

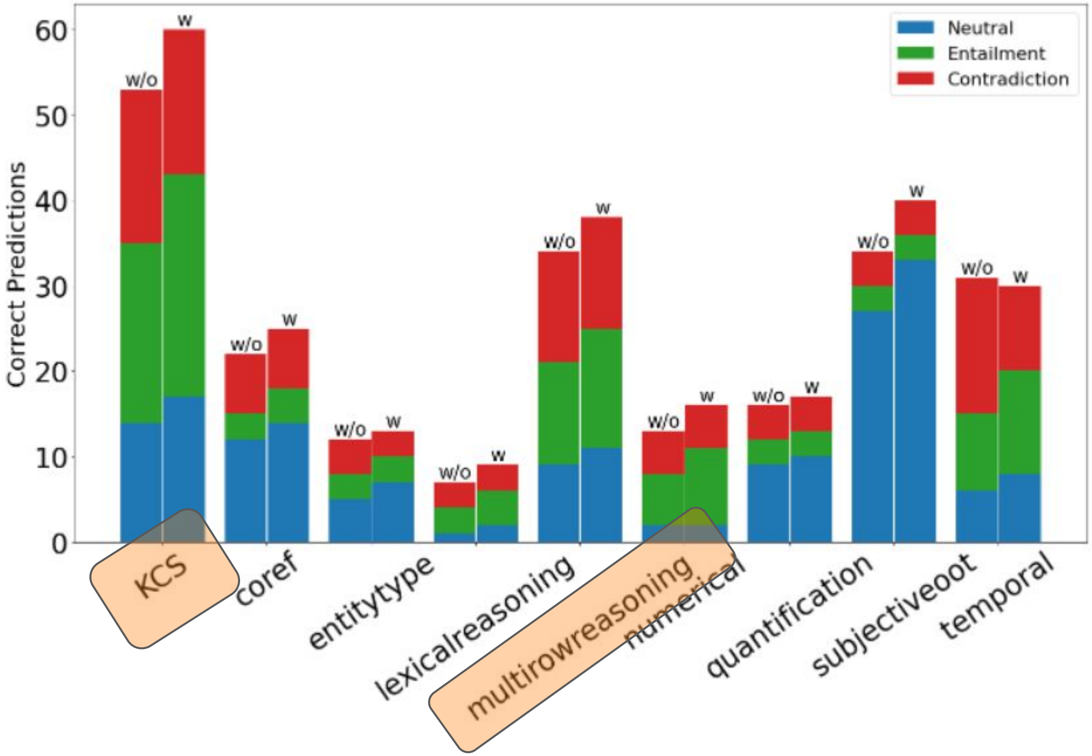


w/o : RoBERTa
w : Trans-KBLSTM

Proposed approach shows improvement almost across all reasoning types!

**Results produced using 3% Training data*

Reasoning Analysis



w/o : RoBERTa
w : Trans-KBLSTM

Let us go through examples for two reasoning types:

1. Knowledge-Common Sense (KCS)
2. Multirow Reasoning

For **lexical reasoning**, refer to paper

***Results produced using 3% Training data*

Knowledge and Common Sense Reasoning

Tables involve ***factual information about world affairs.***

Knowledge Graphs can **supplement this reasoning** abilities to our models.

The ***relation*** between ***kingdom and Monarch*** helps produce correct inference.

Hashemite Kingdom of Jordan Premise	
Premise	The Legislature of Hashemite Kingdom of Jordan are Parliament. The Religion of Hashemite Kingdom of Jordan are 95% Islam (official), 4% Christianity, and 1% Druze, Baha'i. The Government of Hashemite Kingdom of Jordan are Unitary parliamentary constitutional monarchy. The Monarch of Hashemite Kingdom of Jordan is Abdullah II.
Hypothesis	Hashemite Kingdom of Jordan does not have any democracy.
Focused Relation	Kingdom $\xleftrightarrow{IsA}$ Monarch
Gold Label	Contradiction
Prediction	
RoBERTa	Neutral
Trans-KBLSTM	Contradiction

Multi-Row Reasoning

Relational connections encourage *Implicit Extraction*

The relations enforces the model to focus on *right evidence i.e. relevant rows* of the table

Jeff Bridges Premise	
Premise	The Born of Jeff Bridges are December 4, 1949 (age 69) Los Angeles, California, U.S.. The Years active of Jeff Bridges are 1951-present. The Children of Jeff Bridges are 3. The Family of Jeff Bridges are Beau Bridges (brother), and Jordan Bridges (nephew).
Hypothesis	Jeff Bridges started his career as a young child.
Focused Relations	born $\xrightarrow{RelatedTo}$ young born $\xrightarrow{RelatedTo}$ child child $\xrightarrow{RelatedTo}$ age active $\xrightarrow{Co-Hyponym}$ child
Gold Label	Entailment
Prediction	
RoBERTa	Contradiction
Trans-KBLSTM	Entailment

RESULTS AND ANALYSIS

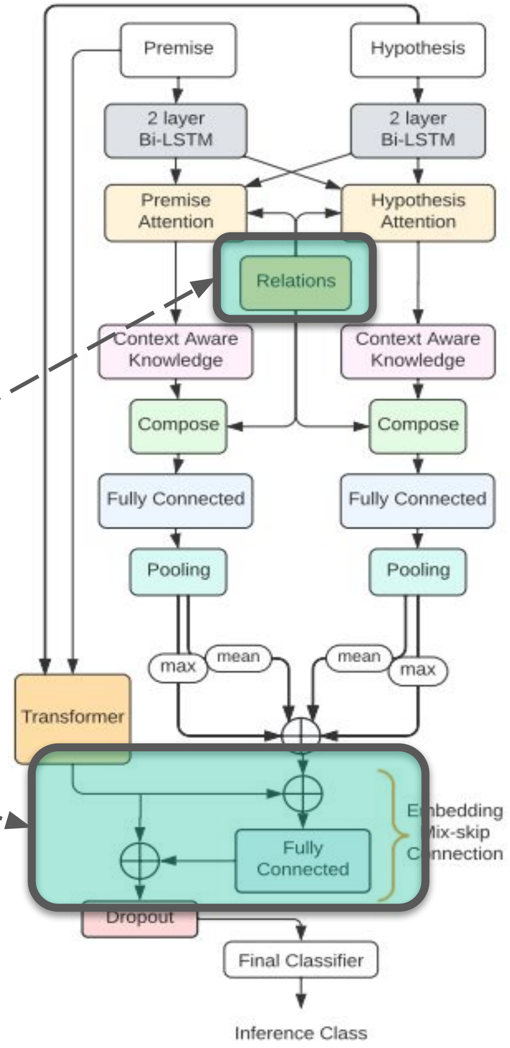
- How effective is our proposed approach for INFOTABS for:
  - *Full Supervision*
  - *Limited Supervision*
- Is our proposed approach *more beneficial* to certain sorts of reasoning types?
- Ablation Study: How important are each individual components?
  - *Embedding mix-skip Connection*
  - *Knowledge Addition*
  - *Independent training of Transformer and LSTM*
  - MNLI Pretraining
  - Transformer Size

refer to paper for details

Embedding Mix-Skip Connection And Knowledge Relations

We ablate these components one by one as follows

1. +Random Noise instead of Knowledge
2. Remove Embedding Skip Connection



Embedding Mix-Skip Connection & Knowledge Relations

Ablations	Dev	α_1	α_2	α_3
Trans-KBLSTM	**67.55**	**65.16**	**64.00**	**63.38**
- Skip Connect	65.72	62.83	60.00	61.55
- KB	60.44	61.88	56.94	55.55
- (KB + Skip Connect)	60.11	61.50	55.94	57.38

Removing Skip Connection

Adding Random noise in place of Knowledge representations

Removing Skip Connection

+Random noise instead of Knowledge representations

Results produced using 1% Training data

Embedding Mix-Skip Connection & Knowledge Relations

Ablations	Dev	α_1	α_2	α_3
Trans-KBLSTM	**67.55**	**65.16**	**64.00**	**63.38**
- Skip Connect	65.72	62.83	60.00	61.55
- KB	60.44	61.88	56.94	55.55
- (KB + Skip Connect)	60.11	61.50	55.94	57.38

Removing Skip Connection

Removing Skip connection and addition of random noise adversely affects model performance

+Random noise in place of Knowledge representations

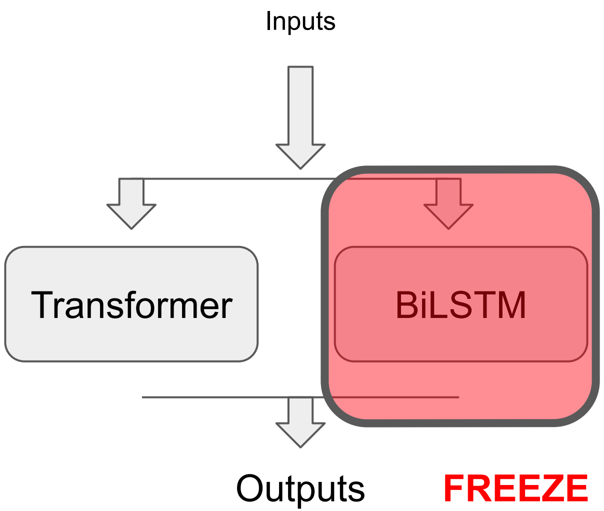
Removing Skip Connection

+Random noise instead of Knowledge representations

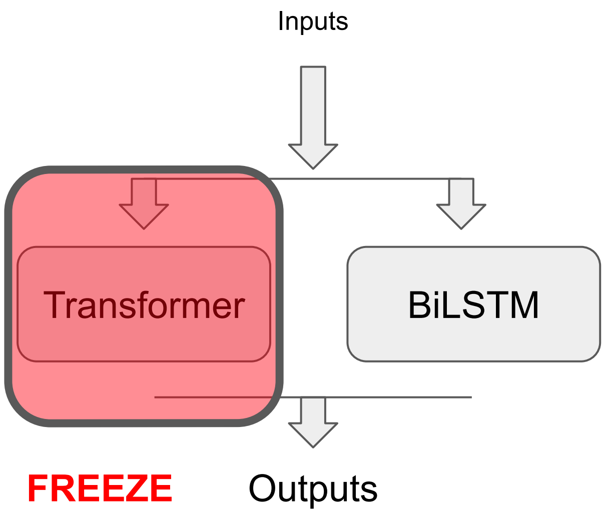
Results produced using 1% Training data

INDEPENDENT TRAINING

STAGE 1



STAGE 2



Independent vs Joint Training

Joint Training better performance!

Ablations	Dev	α_1	α_2	α_3
RoBERTa$_{\text{LARGE}}$	77.30	76.44	70.49	69.05
+ KBLSTM (Independent)	79.22	78.38	71.00	69.22
+ KBLSTM (Joint Train)	**79.92**	**79.62**	**72.10**	**70.21**

Reasons: Brings both embeddings to same representational space

***Results produced using 1% Training data*

TAKEAWAY

This work proposed a novel architecture **Trans-KBLSTM** to solve challenges in ***Knowledge Extraction, Addition and Integration.***

Through **extensive experiments** on the InfoTabS dataset we shown that proposed architecture **enhance inference performance.**

Check out TransKBLSTM: https://trans-kblstm.github.io