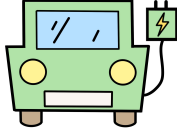


E-KAR  :

A Benchmark for Rationalizing Natural Language Analogical Reasoning

Jiangjie Chen, Rui Xu, Ziquan Fu, Wei Shi, Zhongqiao Li,
Xinbo Zhang, Changzhi Sun, Lei Li, Yanghua Xiao, Hao Zhou



復旦大學
FUDAN UNIVERSITY



ByteDance



Brain Technologies Inc.



華南理工大學
South China University of Technology



UC SANTA BARBARA

Word Analogy Recognition

Query	Q) newton:english
Candidate answers	A) marx:german B) confucius:russian C) caesar:american D) plato:canadian

*An analogical reasoning problem from
The Bigger Analogy Test Set (BATS).*

Word Analogy Recognition

*Word analogy as
multiple-choice QA*

Query	Q) newton:english
Candidate answers	✓ A) marx:german B) confucius:russian C) caesar:american D) plato:canadian

A "word"
or "**term**"



*An analogical reasoning problem from
The Bigger Analogy Test Set (BATS).*

From Linear Analogy to Complex Analogical Reasoning

Linear Analogy

(Ethayarajh et al. 2019)

$$\overrightarrow{\text{king}} - \overrightarrow{\text{man}} + \overrightarrow{\text{woman}} = \overrightarrow{\text{queen}}$$

e.g. *Word2Vec*

From Linear Analogy to Complex Analogical Reasoning

Linear Analogy

(Ethayarajh et al. 2019)

$$\overrightarrow{\text{king}} - \overrightarrow{\text{man}} + \overrightarrow{\text{woman}} = \overrightarrow{\text{queen}}$$

e.g. Word2Vec

Previous work

- *Methods: Hold a connectionist assumption*

From Linear Analogy to Complex Analogical Reasoning

Linear Analogy

(Ethayarajh et al. 2019)

$$\overrightarrow{\text{king}} - \overrightarrow{\text{man}} + \overrightarrow{\text{woman}} = \overrightarrow{\text{queen}}$$

e.g. Word2Vec

Previous work

- *Methods: Hold a connectionist assumption*
- *Benchmarks: Evaluate pre-trained word representations for linear analogy*

Simple Binary Relations

Lexical, morphological, simple semantic relations.



From Linear Analogy to Complex Analogical Reasoning

Linear Analogy

(Ethayarajh et al. 2019)

$$\overrightarrow{\text{king}} - \overrightarrow{\text{man}} + \overrightarrow{\text{woman}} = \overrightarrow{\text{queen}}$$

e.g. Word2Vec

Previous work

- *Methods: Hold a connectionist assumption*
- *Benchmarks: Evaluate pre-trained word representations for linear analogy*

Simple Binary Relations

Lexical, morphological, simple semantic relations.



Not Explainable

Unable to reveal human-like analogical reasoning process.

From Linear Analogy to Complex Analogical Reasoning

Complex Analogy

Q) tea¹:teapot²:teacup³

A) passengers¹:bus²:taxi³

B) magazine¹:bookshelf²:reading room³

C) talents¹:school²:enterprise³

D) textbooks¹:bookstore²:printing factory³

*An analogical reasoning problem from
Civil Service Exams of China.
(Translated)*

The E-KAR Benchmark

Challenging

Sourced from Civil Service Exams of China

Explainable

Free-text Explanations

Bilingual

Chinese & English

Complex Analogy



Chinese

#Problems=1665

#Expl.=5×1665



English

#Problems=1251

#Expl.=5×1251

Dataset	Lang.	Data Size (train / val / test)	# of Terms in Cand.	Has Expl.
SAT	En	0 / 37 / 337	2	✗
Google	En	0 / 50 / 500	2	✗
BATS	En	0 / 199 / 1,799	2	✗
E-KAR 	Zh	1,155 / 165 / 335	2 _(64.5%) , 3 _(35.5%)	✓
	En	870 / 119 / 262	2 _(60.5%) , 3 _(39.5%)	✓

The E-KAR Benchmark

Challenging

Sourced from Civil Service Exams of China

Explainable

Free-text Explanations

Bilingual

Chinese & English

Why are analogical reasoning problems from CSE **challenging**?



Knowledge-intensive term relations

1. Linguistic knowledge
2. Commonsense knowledge
3. Encyclopedic/factual knowledge
4. Cultural knowledge
5. Relations of three terms (35% vs. 0%)
6. Negated facts

Commonsense Knowledge

transport tea
teapot → teacup

Factual Knowledge

cause
Lunar & Solar gravity → Tide

Chinese Idioms:

路见不平:拔刀相助

[Translation] One the road is rough:draw a knife to help

[Meaning] See someone is in trouble:do one's best to help

- husband:job
– Husband is **not** a job.
- car:tires
– A car is **not** made of tires.
– A car consists of tires.



The E-KAR Benchmark

Challenging

*Sourced from Civil
Service Exams of China*

Explainable

Free-text Explanations

Bilingual

Chinese & English

 ***How to Rationalize
Analogical Reasoning?***

Analogical Reasoning: A Psychological Perspective

Structure- mapping theory

(Minnameier et al, 2010)

Q) tea¹:teapot²:teacup³

A) passengers¹:bus²:taxi³

B) magazine¹:bookshelf²:reading room³

C) talents¹:school²:enterprise³

D) textbooks¹:bookstore²:printing factory³

Analogyical Reasoning: A Psychological Perspective

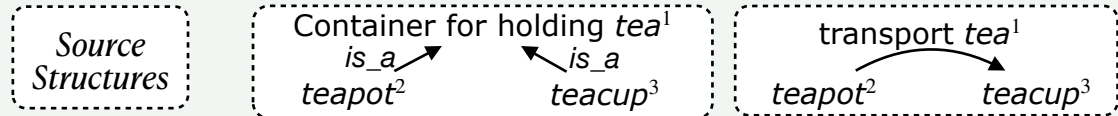
Structure- mapping theory

(Minnameier et al, 2010)

Abduction

Draw a *source structure* that may work for target.

Q) tea¹:teapot²:teacup³



A) passengers¹:bus²:taxi³

B) magazine¹:bookshelf²:reading room³

C) talents¹:school²:enterprise³

D) textbooks¹:bookstore²:printing factory³

Analogyical Reasoning: A Psychological Perspective

Structure- mapping theory

(Minnameier et al, 2010)

Abduction

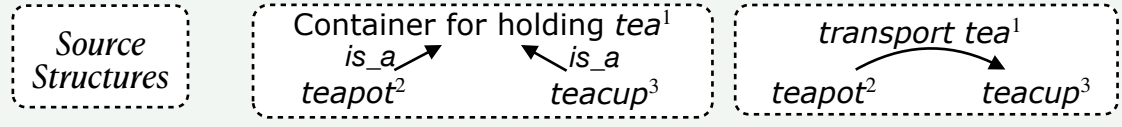
Draw a *source structure* that may work for target.



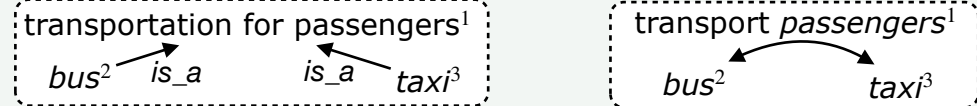
Mapping

Map the structure to the target domain.

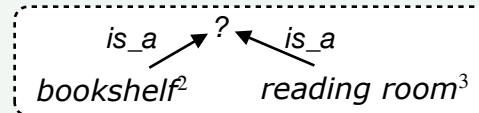
Q) tea¹:teapot²:teacup³



A) passengers¹:bus²:taxi³



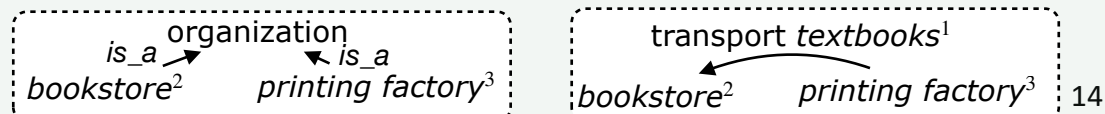
B) magazine¹:bookshelf²:reading room³



C) talents¹:school²:enterprise³



D) textbooks¹:bookstore²:printing factory³



Analogyical Reasoning: A Psychological Perspective

Structure- mapping theory

(Minnameier et al, 2010)

Abduction

Draw a *source structure* that may work for target.

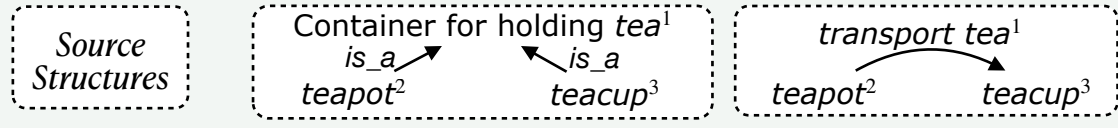
Mapping

Map the structure to the target domain.

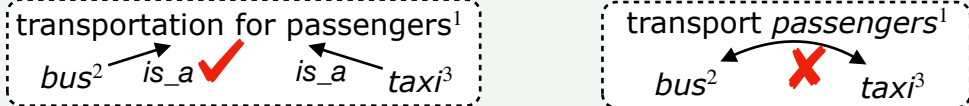
Validation

Validity check and justification w.r.t. solving the target problem.

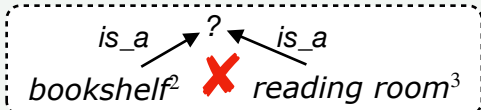
Q) tea¹:teapot²:teacup³



A) passengers¹:bus²:taxi³



B) magazine¹:bookshelf²:reading room³



C) talents¹:school²:enterprise³



D) textbooks¹:bookstore²:printing factory³



How to Rationalize Analogical Reasoning?

🏆 **for Reasoning:**
Being Right for the Right Reasons

Abduction

Mapping

Validation

The Right Reasons:
Verbalize the structure-mapping process into free-text explanations.

Q) tea¹:teapot²:teacup³

Source Structures

Container for holding tea¹
is_a ↗
teapot² ↖ is_a
teacup³

transport tea¹
teapot² → teacup³

Explanation
(free-text)

Both "teapot"² and "teacup"³ are containers for holding "tea"¹. After the "tea"¹ is brewed in the "teapot"², it is transported into the "teacup"³.

A) passengers¹:bus²:taxi³

transportation for passengers¹
bus² is_a ✓ is_a taxi³

transport passengers¹
bus² ✗ taxi³

"Passengers" do not need to be transported into "taxi" after taking a "bus". "Taxi" and "bus" are different ways of transportation.

C) talents¹:school²:enterprise³

organization for talents¹
school² is_a ✓ is_a enterprise³

transport talents¹
school² ✓ enterprise³

Both "school" and "enterprise" are organizations. After "talents" are educated in "school", they are transported into "enterprise".

The E-KAR Benchmark

Challenging

Sourced from Civil Service Exams of China

Explainable

Free-text Explanations

Bilingual

Chinese & English

Example

Refutation

"Passengers" do not need to be transported into "taxi" after taking a "bus". "Taxi" and "bus" are different ways of transportation.

Evidence

**Human-
annotation
Free-text**

*Double-checking
Strategy* for
quality control

(Please check the paper for details.)

The E-KAR Benchmark

Challenging

Sourced from Civil Service Exams of China

Explainable

Free-text Explanations

Bilingual

Chinese & English

Example

Refutation

"Passengers" do not need to be transported into "taxi" after taking a "bus". "Taxi" and "bus" are different ways of transportation.

Evidence

**Human-
annotation
Free-text**

**Explanation
for Every
Query and
Candidate**

*Double-checking
Strategy* for
quality control

Both *Refuting* &
Supporting
Explanation

(Please check the paper for details.)

The E-KAR Benchmark

Challenging

Sourced from Civil Service Exams of China

Explainable

Free-text Explanations

Bilingual

Chinese & English

Example

Refutation

"Passengers" do not need to be transported into "taxi" after taking a "bus". "Taxi" and "bus" are different ways of transportation.

Evidence

**Human-
annotation
Free-text**

**Explanation
for Every
Query and
Candidate**

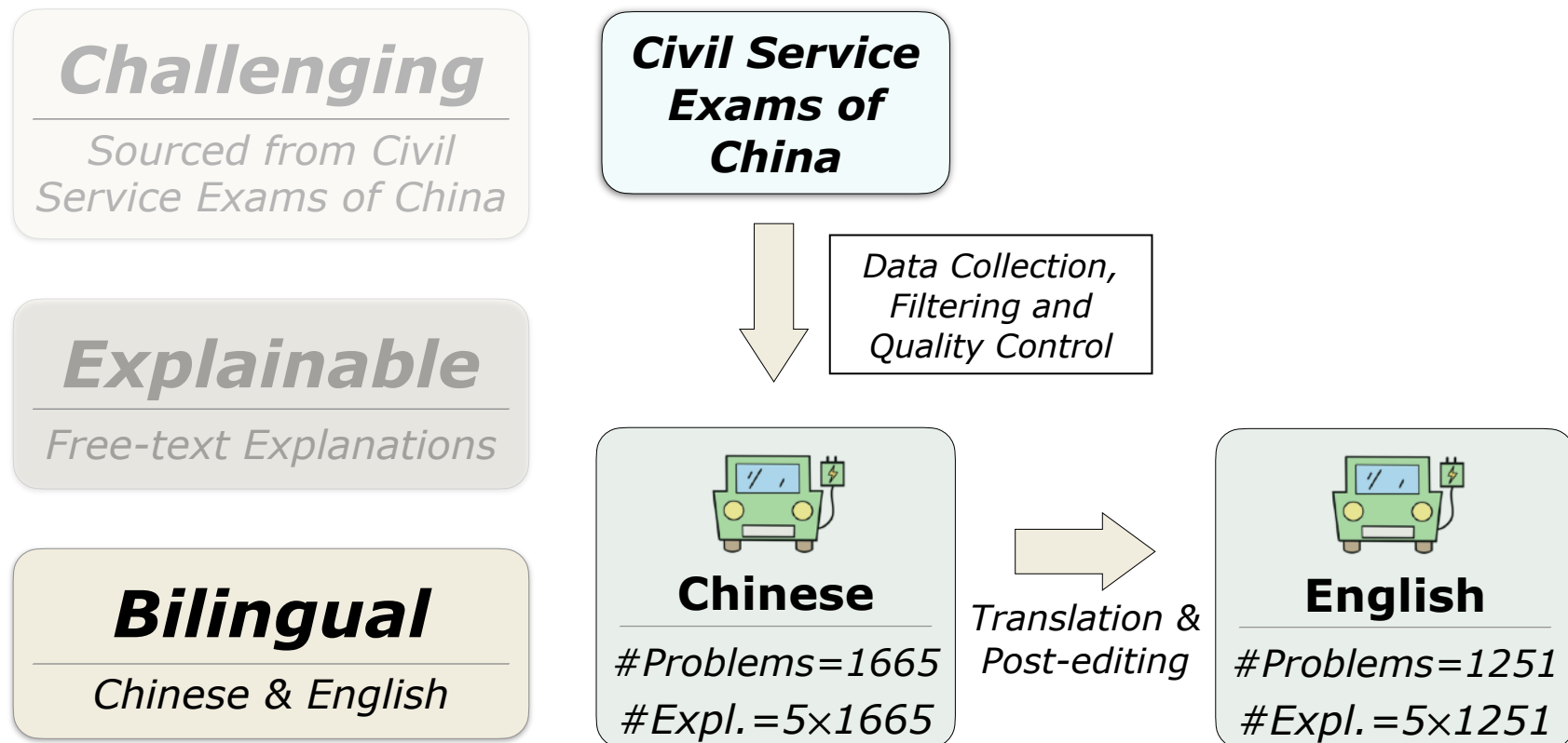
**With Evidence
Showing Why**

*Double-checking
Strategy* for
quality control

Both *Refuting* &
Supporting
Explanation

(Please check the paper for details.)

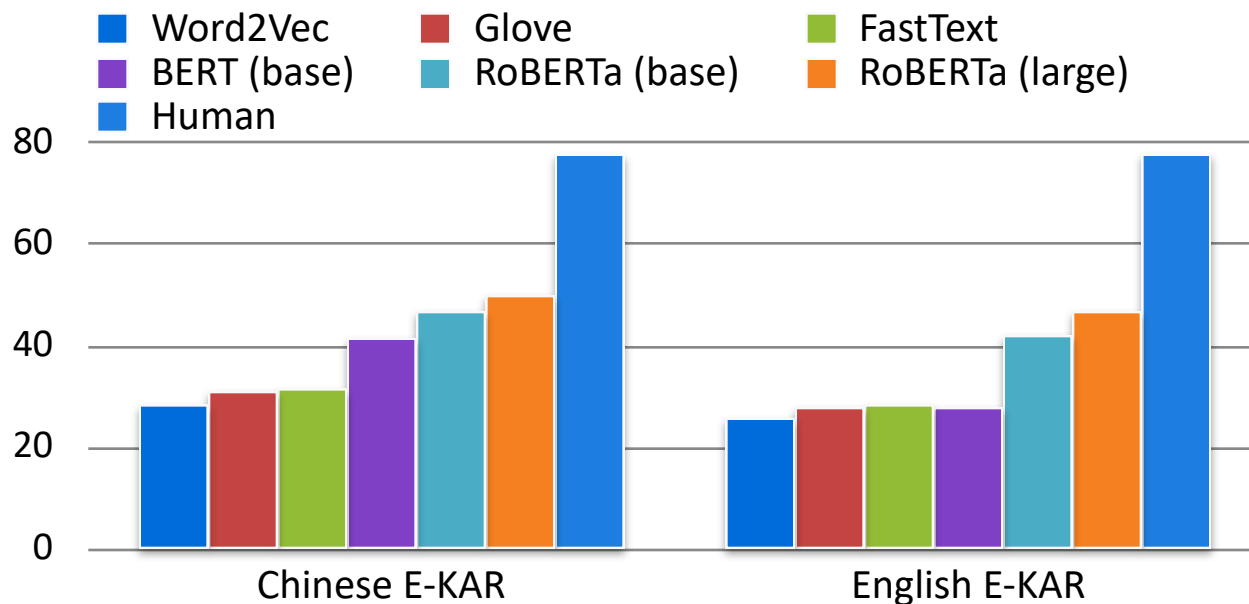
The E-KAR Benchmark



(Please check the paper for details.)

Lessons from Preliminary Exploration of E-KAR

Lesson 1: *W2Vs and LMs both struggle at complex analogical reasoning.*



Humans outperform SOTA models by large margins.

(Please check the paper for details.)

Lessons from Preliminary Exploration of E-KAR

Lesson 2: *LMs struggle at rationalizing analogical reasoning.*

Reminder

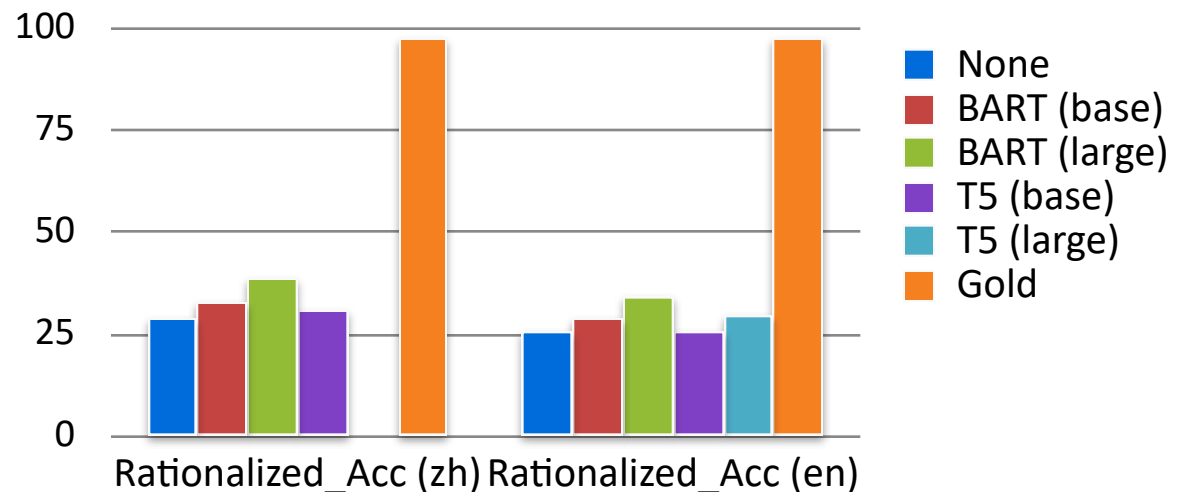


- **Explanation Generation**

- **Input:** Query + Candidates
- **Output:** Free-text explanations for both query $\mathcal{E}_Q$ and candidates $\mathcal{E}_A$

- **Evaluation:** The performance gain in QA when prompted with generated explanations ($\mathcal{E}_Q$ & $\mathcal{E}_A$)

- **Main Metric:** Rationalized QA Accuracy (Acc. with $\mathcal{E}$)
- **QA Model:** RoBERTa-large



1. *Poor quality of generated explanations, improvement over baseline but fall far behind gold.*
2. *Gold explanations can be exploited by Analogical QA models to achieve nearly perfect results (97.7%).*

Lessons from Preliminary Exploration of E-KAR

Lesson 2: *LMs struggle at rationalizing analogical reasoning.*

Error Analysis

1. Unable to generate negated facts for refutation.
2. Generating factually incorrect statements.
3. Biasing towards common patterns.

Ex1. "term1" and "term2" has the same meaning.

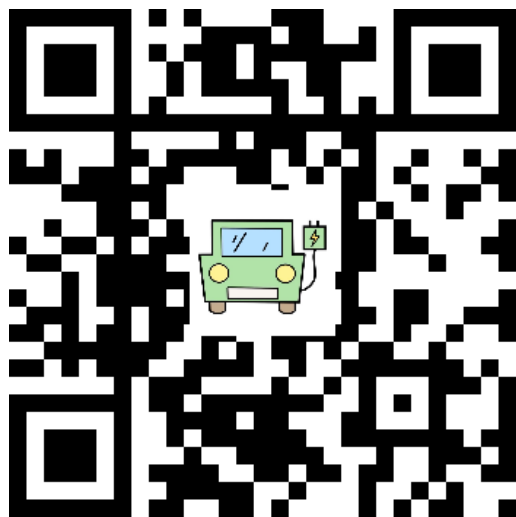
Ex2. "term1" is a "term2".

(Please check the paper for details.)

What is Next?

- ***What we have***: A novel benchmark for rationalizing analogical reasoning, which is ***challenging***, ***explainable*** and ***bilingual***.
- Analogical reasoning by effectively interacting with various kinds of knowledge.
 - e.g. commonsense, factual and *cultural* knowledge.
- Generate reasons with evidence to rationalize reasoning.
 - Particularly, enable models to generate *negated* statements/facts.

Have Fun with E-KAR!



<https://ekar-leaderboard.github.io>



<https://eval.ai/web/challenges/challenge-page/1671/overview>



jjchen19@fudan.edu.cn



<https://jiangjiechen.github.io>