Commonsense Causal Reasoning between Short Texts

Zhiyi Luo, Yuchen Sha , Kenny Q.Zhu, Seung-won Hwang and Zhongyuan Wang

周云晓

Novermber 1, 2017

Outline

- ✓ Background
- Approach
- Experiments
- Conclusions

Task Definition

- What is Commonsense Causal Reasoning ?
  - Capture and understand the causal dependencies amongst events and actions
  - Given a premise and two alternatives, select the more plausible alternative as a cause (or effect) of the premise

Task Definition

- Examples -- 1
  - *Premise*: I fainted this morning.

    What was the *CAUSE* of this?

    ✓ *Alternative 1:* I did not eat for two days.

    Alternative 2: My mobile is out of battery.

Task Definition

- Examples -- 2
  - *Premise*: I knocked on my neighbor's door.

    What happened as an *EFFECT* ?

    ✓ *Alternative 1:* My neighbor invited me in.

    Alternative 2: My neighbor left her house.

Key Challenge

- *Premise*: I knocked on my neighbor's door.

  What happened as an EFFECT?

✓ *Alternative 1:* My neighbor invited me in.

Alternative 2: My neighbor left her house.

Harvest causality knowledge that the action of *knocking* is more likely to cause that of *invitation* than that of *leaving*.

Outline

- Background
- ✓ Approach
- Experiments
- Conclusions

Approach

- Causality Network
- Causal Strength Computation
- Commonsense Causal Reasoning

Causality Network

- Causality can be identified by linguistic patterns known as causal cues

- Examples

(1) The [*storm*]$_{\text{CAUSE}}$ [**caused**]$_{\text{PATTERN}}$ a tremendous amount of [*damage*]$_{\text{EFFECT}}$ on the landing beaches.

(2) The team prepared GIS precipitation and contour maps of the area identifying the [*flooding*]$_{\text{EFFECT}}$ and landslides [**caused by**]$_{\text{PATTERN}}$ the [*rainfall*]$_{\text{CAUSE}}$.

Causality Network

- Causal cues

Table 1: 53 Causal cues. *A* is a cause span, and *B* is an effect span. DET stands for a/an/the/one. BE stands for is/are/was/were.

intra-sentence			inter-sentence		
A lead to B	A leads to B	A led to B	If A, then B	If A, B	B, because A
A leading to B	A give rise to B	A gave rise to B	B because A	B because of A	Because A, B
A given rise to B	A giving rise to B	A induce B	A, thus B	A, therefore B	B, A as a consequence
A inducing B	A induces B	A induced B	Inasmuch as A, B	B, inasmuch as A	In consequence of A, B
A cause B	A causing B	A causes B	B due to A	Due to A, B	B in consequence of A
A caused B	B caused by A	A bring on B	B owing to A	B as a result of A	As a consequence of A, B
A brought on B	A bringing on B	A brings on B	A and hence B	Owing to A, B	B as a consequence of A
B result from A	B resulting from A	B results from A	A, hence B	A, consequently B	A and consequently B
B resulted from A	the reason(s) for/of B BE A		A, for this reason alone , B		
DET effect of A BE B	A BE DET reason(s) of/for B				

- A → a text span that represents the cause
- B → a text span that represents the effect
- Maximum length of A and B → 10

Causality Network

- Causal pair (i_c , j_e)
  - $i_c \rightarrow$ term i in A $j_e \rightarrow$ term j in B

Example:

The storm **caused** *a tremendous amount of damage on the landing beaches*.

(storm$_c$, tremendous$_e$), (storm$_c$, amount$_e$), (storm$_c$, damage$_e$), (storm$_c$, landing$_e$), (storm$_c$, beach$_e$)

Causality Network

- Directed network of causal relations

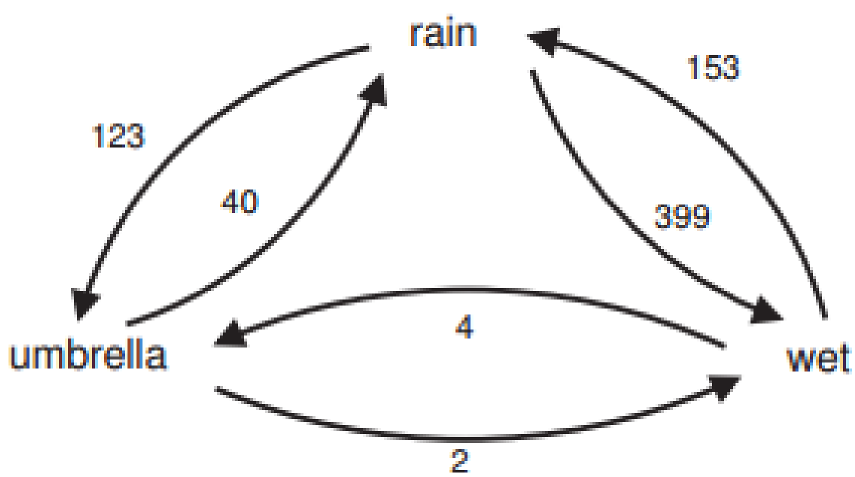


Figure 1: A fragment of causal network

- Node:
  lemmatized term
- Directed edge:
  causal relation
- numbers:
  causality
  co-occurrences

Causal Strength Computation

- Necessity causality
  - Cause *i* must be present in order for the effect *j* to take place → （rainfall_c, flooding_e）
  - $p(i_c | j_e)$
- Sufficiency causality
  - Cause *i* is all it takes bring about the effect *j* → (storm_c, damage_e)
  - $p(j_e | i_c)$

Causal Strength Computation

- Necessity causality strength

$$CS_{nec}(i_c, j_e) = \frac{p(i_c|j_e)}{p^{\alpha}(i_c)} = \frac{p(i_c, j_e)}{p^{\alpha}(i_c)p(j_e)}$$

$$p(i_c) = \frac{\sum_{w \in W} f(i_c, w_e)}{M}$$

$$p(j_e) = \frac{\sum_{w \in W} f(w_c, j_e)}{M}$$

$$p(i_c, j_e) = \frac{f(i_c, j_e)}{N}$$

- Sufficiency causality strength

$$CS_{suf}(i_c, j_e) = \frac{p(j_e|i_c)}{p^{\alpha}(j_e)} = \frac{p(i_c, j_e)}{p(i_c)p^{\alpha}(j_e)},$$

Causal Strength Computation

- Causality strength computation

$$CS(i_c, j_e) = CS_{nec}(i_c, j_e)^{\lambda} CS_{suf}(i_c, j_e)^{1-\lambda}$$

- Compute the causal strength between every pair of terms in the causality network

Commonsense Causal Reasoning

- Goal :

Compute whether alternative *a1* or *a*2 is more plausible with respect to the premise *p*

- Compute the causal strength from text $T1$ to text $T2$:

$$CS_T(T_1, T_2) = \frac{1}{|T_1| + |T_2|} \sum_{i \in T_1} \sum_{j \in T_2} CS(i, j)$$

Commonsense Causal Reasoning

- Compare the overall causality strength $CS_T\,(p, a1)$ and $CS_T\,(p, a2)$, assuming p is asking for an *effect*.

- Compare the overall causality strength $CS_T\,(a1, p)$ and $CS_T\,(a2, p)$, assuming p is asking for an *cause*.

Outline

- Background
- Approach
- ✓ Experiments
- Conclusions

Experiments

- Extraction of Causality Network
  - A large web text corpus (a snapshot of Bing, 10TB, 1.6billion web pages)
  - The number of unique lemmatized terms (nodes) is 64,436
  - The average frequency of pairs is 10.54

Experiments

- COPA task dataset
  - Choice of Plausible Alternatives
  - Development question set → 500
  - Test question set → 500

    - *Premise*: I knocked on my neighbor's door.

      What happened as an EFFECT?

      ✓ *Alternative 1:* My neighbor invited me in.

      Alternative 2: My neighbor left her house.

Experiments

- COPA results comparison

Data Source	Methods	Accuracy(%)
Web corpus	PMI (W=5)	61.6%
	PMI (W=10)	61.0%
	PMI (W=15)	60.4%
	PMI (W=25)	61.2%
	$CS_{\lambda=0.5}$	**64.8%**
Gutenberg	PMI (W=5)	58.8%
	PMI (W=25)	58.6%
LDC Gigaword	UTDHLT Bigram PMI	61.8%
	UTDHLT SVM	63.4%
ConceptNet	Fuzzy match	51.3%
1-Million Stories	PMI (W=25)	65.2%
10-Million Stories	PMI (W=25)	**65.4%**
CausaNet	$CS_{\lambda=1.0}$	**70.2** %

Experiments

- COPA results comparison

Data Source	Methods	Accuracy(%)
Web corpus	PMI (W=5)	61.6%
	PMI (W=10)	61.0%
	PMI (W=15)	60.4%
	PMI (W=25)	61.2%
	$CS_{\lambda=0.5}$	**64.8%**
Gutenberg	PMI (W=5)	58.8%
	PMI (W=25)	58.6%
LDC Gigaword	UTDHLT Bigram PMI	61.8%
	UTDHLT SVM	63.4%
ConceptNet	Fuzzy match	51.3%
1-Million Stories	PMI (W=25)	65.2%
10-Million Stories	PMI (W=25)	**65.4%**
CausaNet	$CS_{\lambda=1.0}$	**70.2** %

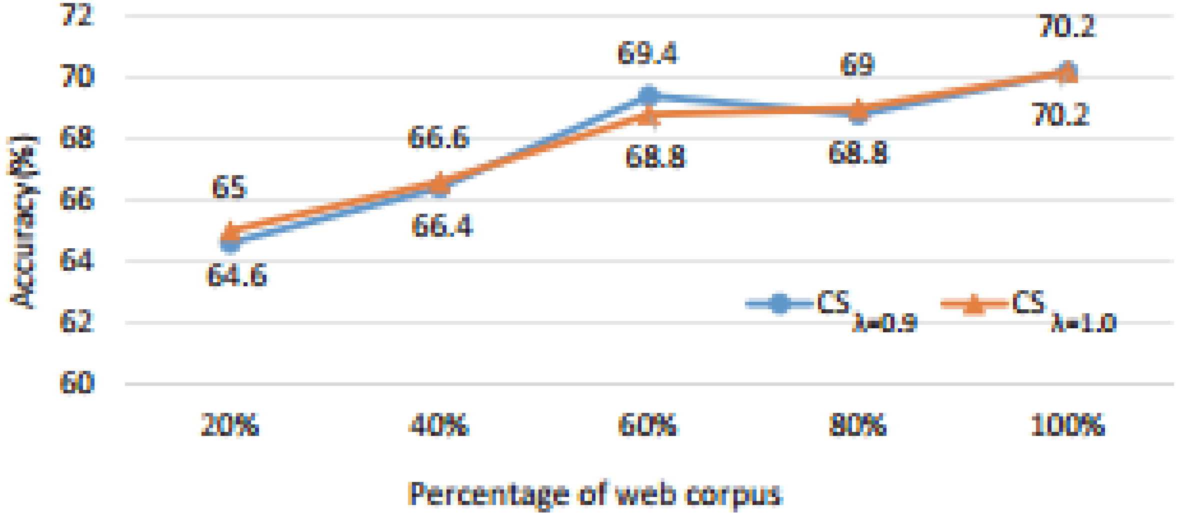
Experiments

- COPA results comparison

Data Source	Methods	Accuracy(%)
Web corpus	PMI (W=5)	61.6%
	PMI (W=10)	61.0%
	PMI (W=15)	60.4%
	PMI (W=25)	61.2%
	$CS_{\lambda=0.5}$	**64.8%**
Gutenberg	PMI (W=5)	58.8%
	PMI (W=25)	58.6%
LDC Gigaword	UTDHLT Bigram PMI	61.8%
	UTDHLT SVM	63.4%
ConceptNet	Fuzzy match	51.3%
1-Million Stories	PMI (W=25)	65.2%
10-Million Stories	PMI (W=25)	**65.4%**
CausaNet	$CS_{\lambda=1.0}$	**70.2 %**

Experiments

- Effect of web data size



Experiments

- Effect of λ

Data Source	Methods	Accuracy(%)
Web corpus	$p(j_e\|i_c)$	58.2%
	$p(i_c\|j_e)$	62.8%
	$CS_{\lambda=0.5}$	**64.8%**
	$CS_{\lambda=0.7}$	63.4%
	$CS_{\lambda=0.9}$	63.0%
	$CS_{\lambda=1.0}$	63.0%
CausalNet	$p(j_e\|i_c)$	56.2%
	$p(i_c\|j_e)$	60.2%
	$CS_{\lambda=0.5}$	68.8%
	$CS_{\lambda=0.7}$	69.4%
	$CS_{\lambda=0.9}$	**70.2%**
	$CS_{\lambda=1.0}$	**70.2%**

Conclusion

- Automatically construct graph-based representation of a causality knowledge base
- Utilize both sufficiency and necessary causality evidences to model the causality strength between terms
- Outperform the previous best approach for solving COPA task

Thank you!