NOV 2020

DSO NLP Reading Group
Presented by Eugene

On Symbolic and Neural Commonsense Knowledge Graphs

Hwang, J. D., Bhagavatula, C., Bras, R. L., Da, J., Sakaguchi, K., Bosselut, A., & Choi, Y.

CONTRIBUTIONS

- ATOMIC 2020, a **commonsense knowledge graph (CSKG)** made up of **1.33M tuples** describing social, physical, and event-centric aspects of everyday inferential knowledge.

#Crowdsourced | #>1MTriples | NEW ObjectUse | NEW HinderedBy | #Symbolic

- Show that **ATOMIC 2020** is **more accurate** than SOTA CSKGs ATOMIC, ConceptNet and TransOMCS.

#Crowdsourced | #>91Percent | #Symbolic

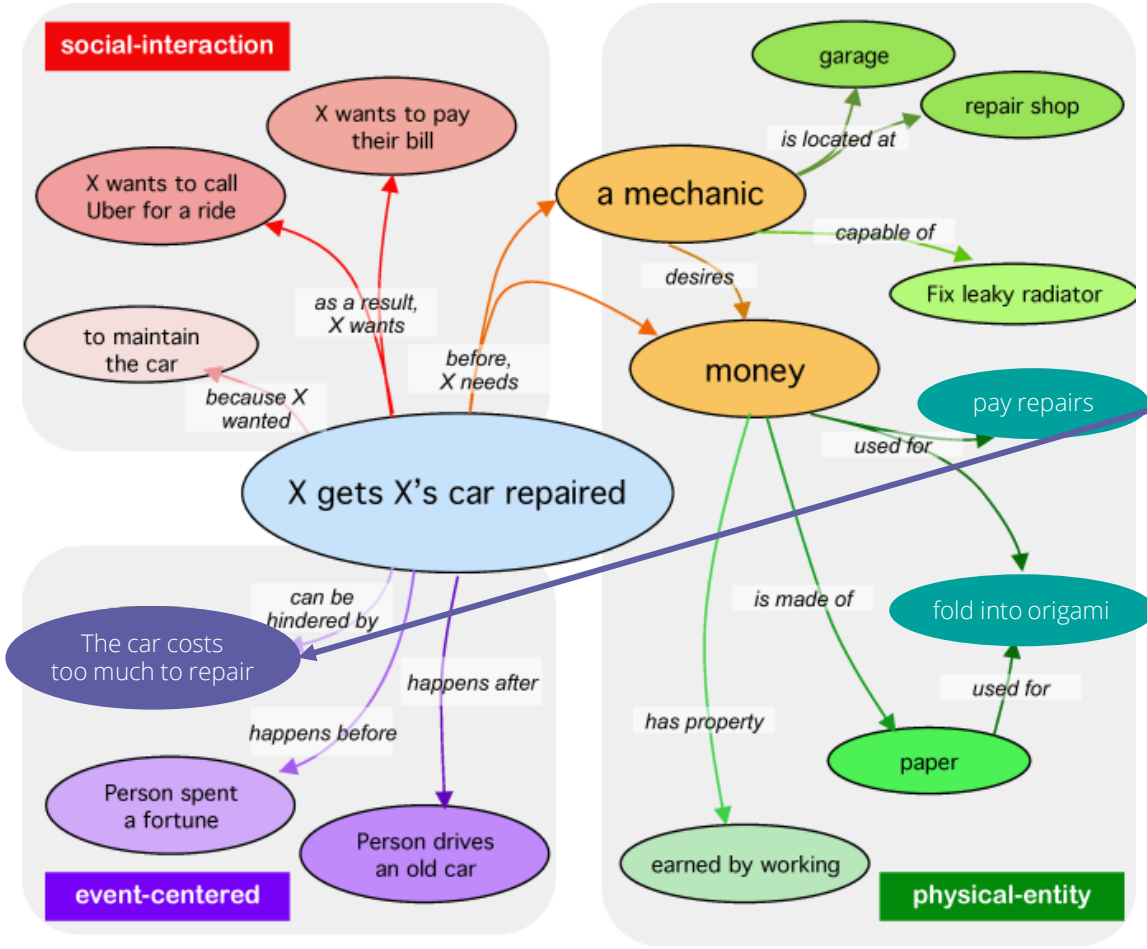
  - A coverage assesment showed ATOMIC 2020 covers all of ATOMIC and half of ConceptNet. ATOMIC covers only 60% of ATOMIC 2020 and the other two cover from 1-10% of ATOMIC 2020.

- Show that on a task where a language model hypothesizes knowledge tuples (**predict tail given head and relation**), **transferring knowledge from a high-quality symbolic KG performs better** than few-shot learning with **GPT3** with much fewer parameters.

#HumanEval | #BLEU, #METEOR, #ROUGE, #CIDEr, #BERTScore | #~12PercentBetter | #Neural

ATOMIC 2020

A Symbolic Commonsense Knowledge Graph

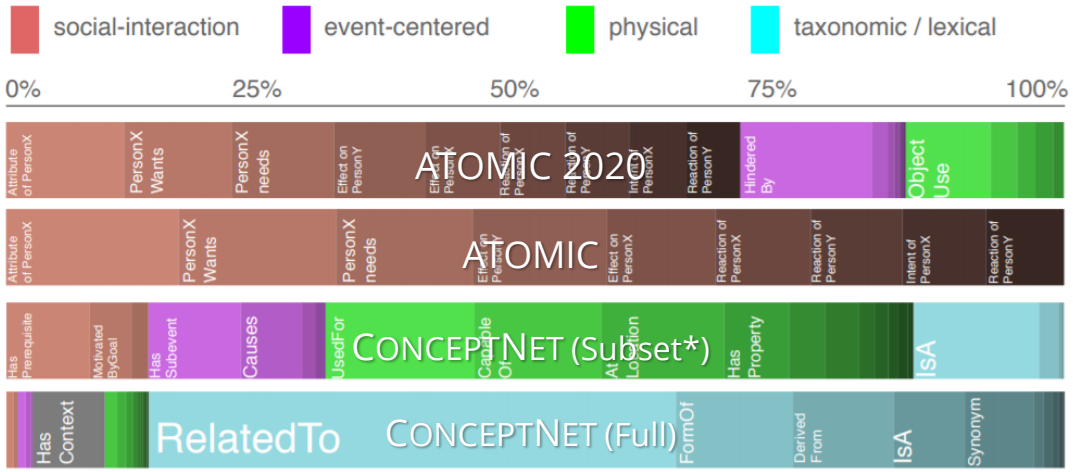


	Head	Relation	Tail	Size
Physical-Entity	bread	ObjectUse NEW	make french toast	165,590
		AtLocation*	basket; pantry	20,221
		MadeUpOf	dough; wheat	3,345
		HasProperty*	cooked; nice to eat	5,617
	baker	CapableOf*	coat cake with icing	7,968
		Desires*	quality ingredients	2,737
		Not Desires*	bad yeast	2,838
Event-Centered	X runs out of steam	IsAfter	X exercises in the gym	22,453
		HasSubEvent	become tired	12,845
		IsBefore	X hits the showers	23,208
		HinderedBy NEW	drinks too much coffee	106,658
		Causes	takes a break	376
		xReason	did not eat breakfast	334
	X watches ___ anyway	isFilledBy	the game; the TV	33,266
Social Interaction	X runs out of steam	xNeed	do something tiring	128,955
		xAttr	old; lazy; lethargic	148,194
		xEffect	drinks some water	115,124
		xReact	tired	81,397
		xWant	to get some energy	135,360
	X votes for Y	xIntent	to give support	72,677
		oEffect	receives praise	80,166
		oReact	grateful; confident	67,236
		oWant	thank X; celebrate	94,548

* - Relations that reflect semantically identical categories to CONCEPTNET

COMPARING SYMBOLIC CSKGS

COMPOSITION



- * - The CONCEPTNET subset consists of only common sense relations, excluding lexical (e.g., Synonym, RelatedTo) knowledge.
- The authors did not include the composition of the 18.48M TransOMCS [4] that is automatically converted from syntactic parses of sentences from various web sources including Wikipedia, Yelp, and Reddit.

COVERAGE

Coverage Precision - Average number of times (%) tuple in Source KB is found in Target KB.

ATOMIC 2020 covers all of ATOMIC and ~half of ConceptNet.

SOURCE KB \ TARGET KB	ATOMIC	CN	T-OMCS	ATOMIC2020
ATOMIC	-	0.1	0.0	100.0
CONCEPTNET	0.3	-	5.5	45.6
TRANSOMCS	0.0	0.4	-	0.3
ATOMIC2020	60.2	9.3	1.4	-

ATOMIC covers only 60.2% of ATOMIC 2020 and the other two cover from 9.3% and 1.4% of ATOMIC 2020 respectively.

ACCURACY

- 3,000 random instances were extracted from each of the knowledge graphs for a crowdsourced evaluation of the tuples.

CSKG	Accept	Reject	No Judgement
Atomic2020	**91.3**	**6.5**	2.2
Atomic	88.5	10.0	1.5
ConceptNet	88.6	7.5	3.9
TransOMCS	41.7	53.4	4.9

ATOMIC2020	ATOMIC	Relation	CN	T-OMCS
92.3		AtLocation*	89.4	34.3
93.9		CapableOf*	84.4	50.0
94.6		Causes	90.0	50.0
96.9		Desires*	96.3	48.2
93.9		HasProperty*	86.3	52.4
82.3	NEW	ObjUse/UsedFor	**96.3**	31.6
98.5		NotDesires*	96.3	
96.9		HasSubevent	88.1	57.7
75.4		MadeUpOf/MadeOf	**88.1**	15.9
96.9	NEW	HinderedBy		
93.1	89.9	xIntent/MotivByGoal	84.4	27.1
82.3	88.4	xWant/CausesDesire	**90.0**	35.9

COMET

Commonsense Transformers for Automatic Knowledge Graph Construction [3]

- Transfer implicit knowledge from deep pre-trained language models to generate (hypothesize) explicit knowledge in CSKGs.

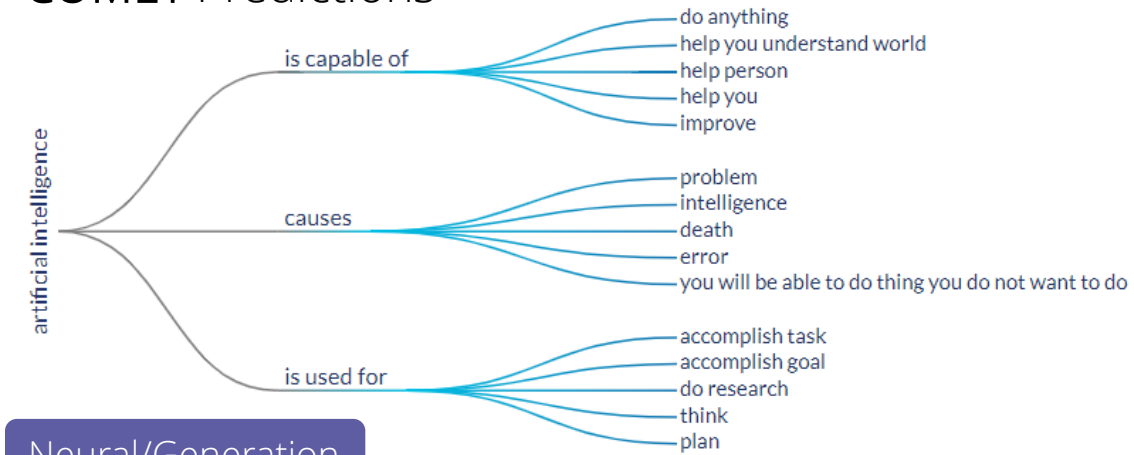
ConceptNet V5.8 CSKG

artificial intelligence

- is used for → understanding human mind and brains
- is defined as → fake , not aquired
- Is a type of → computer science

Symbolic

COMET Predictions



Neural/Generation

Training

`<head> <relation> [GEN] <tail> [SEP]`

FINE TUNING

Pre-trained Language Model

GPT2-XL | BART | GPT

Inference

`<head> <relation> [GEN]`

COMET

GPT2-XL | BART | GPT

`… [GEN] <tail> [SEP]`

COMPARING NEURAL CSKGS

		Accept	Reject	No Judgm.
ATOMIC2020	GPT2-XL	36.6	62.5	0.9
	GPT-3	73.0	24.6	2.5
	COMET(GPT2-XL)	72.5	26.6	0.9
	COMET(BART)	**84.5**	**13.8**	1.7
ATOMIC	GPT2-XL	38.3	61.2	0.4
	COMET(GPT2-XL)	64.1	34.7	1.2
	COMET(BART)	**83.1**	**15.3**	1.6
CONCEPTNET	GPT2-XL	50.3	42.1	7.7
	COMET(GPT2-XL)	74.5	19.0	6.4
	COMET(BART)	**75.5**	**17.9**	6.6
TRANSOMCS	GPT2-XL	**28.7**	**53.5**	17.8
	COMET(GPT2-XL)	26.9	60.9	12.2
	COMET(BART)	23.8	65.9	10.3

		Bleu-1	Bleu-2	Bleu-3	Bleu-4	METEOR	ROUGE-L	CIDEr	BERT Score
ATOMIC2020	GPT2-XL	0.101	0.028	0.010	0.003	0.082	0.098	0.047	0.395
	GPT-3	0.299	0.153	0.081	0.048	0.182	0.255	0.175	0.540
	COMET(GPT2-XL)	0.407	0.248	0.171	0.124	0.292	0.485	0.653	0.638
	COMET(BART)	**0.469**	**0.286**	**0.189**	**0.130**	**0.330**	**0.495**	**0.658**	**0.639**
ATOMIC	GPT2-XL	0.083	0.029	0.011	0.005	0.081	0.087	0.045	0.386
	COMET(GPT2-XL)	0.419	0.296	**0.228**	**0.189**	0.292	0.517	0.733	0.634
	COMET(BART)	**0.515**	**0.324**	0.220	0.159	**0.347**	**0.546**	**0.740**	**0.646**
CONCEPTNET	GPT2-XL	0.044	0.012	0.004	0.002	0.064	0.057	0.050	0.389
	COMET(GPT2-XL)	0.155	**0.119**	**0.095**	**0.078**	**0.134**	**0.193**	**0.425**	**0.552**
	COMET(BART)	**0.172**	0.111	0.072	0.049	0.130	0.184	0.368	0.535
TRANSOMCS	GPT2-XL	0.028	0.001	0.000	0.000	0.093	0.053	0.013	0.351
	COMET(GPT2-XL)	0.301	0.000	0.000	0.000	0.180	0.302	0.254	0.677
	COMET(BART)	**0.351**	**0.170**	**0.003**	0.000	**0.198**	**0.352**	**0.297**	**0.678**

CALL FOR COLLABORATION

Downstream Use Cases for COMET and CSKGs

Conversational / Replies

RAG: Retrieval-Augmented Generation [8]

\+

COMET

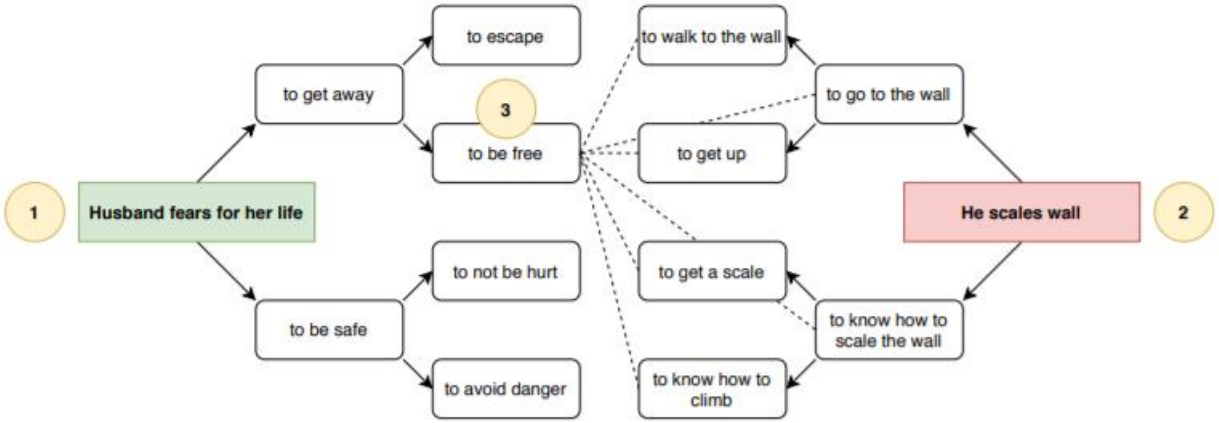
ATOMIC

ConceptNet

\+

DialogRPT [9]

Narrative



Plot Graph Generation (C2PO) [7]

Sarcastic Re-writing (R^3) [5]

Literal Input 1	I hate getting sick from fast food.
GenSarc1	I love getting sick from fast food.
GenSarc2	[I love getting sick from fast food.] [Stomach ache is just an additional side effect.]
Human 1	Shout out to the Mc donalds for giving me bad food and making me sick right before work in two hours.
Literal Input 2	I inherited unfavorable genes from my mother.
GenSarc3	I inherited great genes from my mother.
GenSarc4	[I inherited great genes from my mother.] [Ugly goes down to the bone.]
Human 2	Great I inherited all of my mother's GOOD genes

Mystery Story Generation (C2PO) [7]

Holmes decides go. Holmes wants to go. Holmes begins to see something. Holmes begins to look around. Holmes notices a pair of trouser knees. Holmes wants to clean up. Holmes begins take a shower. Holmes wants to get ready. Holmes wants to walk to the store. Holmes taps in front of Wilson's shop. Holmes tries say hello. Holmes wants start the car. Holmes tries to drive to the scene. He calls Police Inspector Jones.

REFERENCES

1. Hwang, J. D., Bhagavatula, C., Bras, R. L., Da, J., Sakaguchi, K., Bosselut, A., & Choi, Y. (2020). COMET-ATOMIC 2020: On Symbolic and Neural Commonsense Knowledge Graphs. arXiv preprint arXiv:2010.05953.
2. Bosselut, A., Rashkin, H., Sap, M., Malaviya, C., Celikyilmaz, A., & Choi, Y. (2019). COMET: Commonsense transformers for automatic knowledge graph construction. In ACL 2019.
3. Sap, M., Le Bras, R., Allaway, E., Bhagavatula, C., Lourie, N., Rashkin, H., & Choi, Y. (2019). Atomic: An atlas of machine commonsense for if-then reasoning. In AAAI (Vol. 33, pp. 3027-3035).
4. H. Zhang, D. Khashabi, Y. Song, and D. Roth (2020) TransOMCS: from linguistic graphs to commonsense knowledge. In IJCAI 2020.
5. T. Chakrabarty, D. Ghosh, S. Muresan, and N. Peng (2020) R^3: reverse, retrieve, and rank for sarcasm generation with commonsense knowledge. In ACL 2020.
6. W. R. Kearns, N. Kaura, M. Divina, C. V. Vo, D. Si, T. M. Ward, and W. Yuwen (2020) A wizard-of-oz interface and persona-based methodology for collecting health counseling dialog. Extended Abstracts of the 2020 CHI Conference on Human Factors in Computing Systems.
7. P. Ammanabrolu, W. Cheung, W. Broniec, and M. Riedl (2020) Automated storytelling via causal, commonsense plot ordering. arXiv preprint arXiv:2009.00829.
8. Lewis, P., Perez, E., Piktus, A., Petroni, F., Karpukhin, V., Goyal, N., and Riedel, S. (2020). Retrieval-augmented generation for knowledge-intensive NLP tasks. arXiv preprint arXiv:2005.11401.
9. Gao, X., Zhang, Y., Galley, M., Brockett, C., and Dolan, B. (2020). Dialogue Response Ranking Training with Large-Scale Human Feedback Data. In EMNLP 2020.

LINKS

ATOMIC https://homes.cs.washington.edu/~msap/atomic/

COMET Demo https://mosaickg.apps.allenai.org/

COMET Source Code https://github.com/atcbosselut/comet-commonsense

These Slides https://github.com/eugenesiow/my-slides/blob/master/talks/comet-atomic-2020.pdf

https://gitly.hopto.org/eugene/awesome-research