COMET : Commonsense Transformers for Automatic Knowledge Graph Construction

Antoine Bosselut, Hannah Rashkin, Maarten Sap,
Chaitanya Malaviya, Asli Celikyilmaz, Yejin Choi

ACL 2019'

Paul G. Allen School of Computer Science & Engineering, WA, USA
Allen Institute for Artificial Intelligence, WA, USA
Microsoft Research, Redmond, WA, USA

Presented by Dong Hui Im
ehdgnl101@korea.ac.kr
Data Intelligence Laboratory, Korea University
26th March 2021

Problem Statement

How to construct commonsense knowledge graph automatically?

Previous work

[ConceptNet]

en **bicycle**

An English term in ConceptNet 5.5

Sources: Open Mind Common Sense contributors, DBPedia 2015, JMDict 1.07, OpenCyc 2012, German Wiktionary, English Wiktionary, French Wiktionary, and Open Multilingual WordNet

Synonyms

- tr bisiklet →
- en wheel (n) →
- ja 銀輪 (n) →
- it bici →
- ar دَرّاجة هَوائِيّة (n) →
- fr vélo →
- en cycle →
- da cykel (n) →
- it bicicletta →

Related terms

- en biker (n) →
- fr bécane →
- en tricycle →
- en penny farthing (n) →
- br marc'h houarn (n) →
- en propel →
- ca bicicleta (n) →
- en like riding bicycle →
- ee gasɔ (n) →

bicycle is a type of…

- en a two wheel vehicle →
- en means of transportation →
- en a machine →
- en ride (v) →
- en an efficient form of human transportation →
- en toy →
- en transportation →
- en wheeled vehicle (n) →

bicycle is used for…

- en transportation →
- en riding →
- en Racing →
- en personal transport →
- en ride (v) →
- en travelling on →
- en rush (v) →
- en cause cultural change →
- en traveling →

Knowledge graph that connects multilingual words and phrases.

Previous work

[ATOMIC]

Natural language commonsense graph using logical structure.

Contribution

1. They develop a generative approach to knowledge base construction.

2. They develop a framework for using large-scale transformer language models to learn to produce commonsense knowledge tuples.

3. They perform an empirical study on the quality, novelty, and diversity of the commonsense knowledge produced by our approach for two domains, ATOMIC and ConceptNet.

Task

Subject + Relation

↓

generate Object

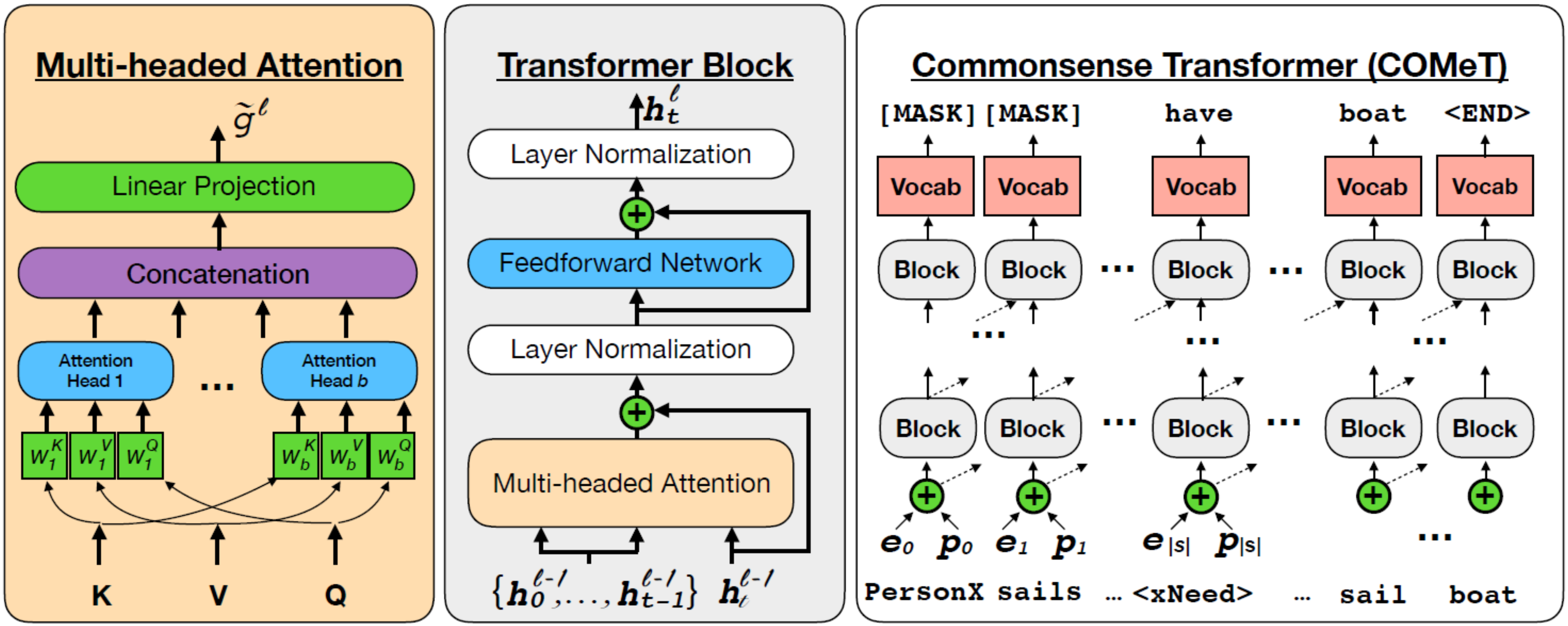
S = take a nap

R = Causes

O = have energy

Model

[Overview]

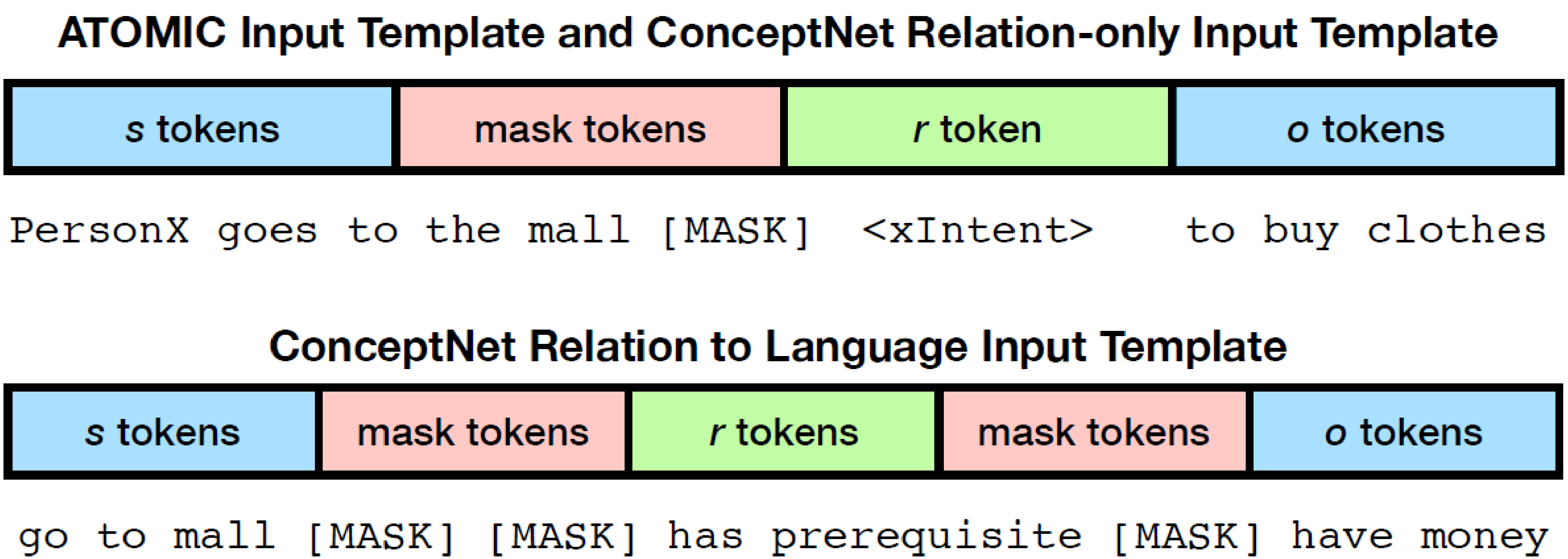


(a) (b) (c)

Using GPT single stream transformer model.

Model

[Input Encoder]



For each dataset, using different encoding style.

Q & A

Code & Setting

[Code] Publicly available

https://github.com/atcbosselut/comet-commonsense

[Setting]

GPT-base model

- 12 layers, 768 dimensional hidden states, 12 attention heads
- Dropout 0.1, Using GeLU
- Batch size64

ATOMIC

- Warmup with 100 minibatches Learning rate : 6.25e-5
- Learning rate decay and early stopping

ConceptNet

- Warmup with 200 minibatches Learning rate : 1e-5

Details – Appendix on paper

Experiments

[ATOMIC]

Model	PPL[5]	BLEU-2	N/T *sro*[6]	N/T *o*	N/U *o*
9ENC9DEC (Sap et al., 2019)	-	10.01	100.00	8.61	40.77
NearestNeighbor (Sap et al., 2019)	-	6.61	-	-	-
Event2(IN)VOLUN (Sap et al., 2019)	-	9.67	100.00	9.52	45.06
Event2PERSONX/Y (Sap et al., 2019)	-	9.24	100.00	8.22	41.66
Event2PRE/POST (Sap et al., 2019)	-	9.93	100.00	7.38	41.99
COMET (- pretrain)	15.42	13.88	100.00	7.25	45.71
COMET	**11.14**	**15.10**	100.00	**9.71**	**51.20**

Model	oEffect	oReact	oWant	xAttr	xEffect	xIntent	xNeed	xReact	xWant	Avg
9Enc9Dec (Sap et al., 2019)	22.92	32.92	35.50	52.20	47.52	51.70	48.74	63.57	51.56	45.32
Event2(In)voluntary (Sap et al., 2019)	26.46	36.04	34.70	52.58	46.76	61.32	49.82	71.22	52.44	47.93
Event2PersonX/Y (Sap et al., 2019)	24.72	33.80	35.08	52.98	48.86	53.93	54.05	66.42	54.04	46.41
Event2Pre/Post (Sap et al., 2019)	26.26	34.48	35.78	52.20	46.78	57.77	47.94	72.22	47.94	46.76
COMET (- pretrain)	25.90	35.40	40.76	48.04	47.20	58.88	59.16	64.52	65.66	49.50
COMET	**29.02**	**37.68**	**44.48**	**57.48**	**55.50**	**68.32**	**64.24**	**76.18**	**75.16**	**56.45**

Experiments

[ATOMIC]

COMET Decoding method	oEffect	oReact	oWant	xAttr	xEffect	xIntent	xNeed	xReact	xWant	Avg
Top-5 random sampling (n=2500 per relation)	34.60	44.04	35.56	64.56	55.68	58.84	46.68	80.96	58.52	53.27
Top-10 random sampling (n=5000 per relation)	25.20	37.42	27.34	49.20	47.34	47.06	38.24	72.60	48.10	43.61
Beam search - 2 beams (n=1000 per relation)	43.70	54.20	47.60	**84.00**	51.10	73.80	50.70	85.80	78.70	63.29
Beam search - 5 beams (n=2500 per relation)	37.12	45.36	42.04	63.64	**61.76**	63.60	57.60	78.64	68.40	57.57
Beam search - 10 beams (n=5000 per relation)	29.02	37.68	44.48	57.48	55.50	68.32	64.24	76.18	75.16	56.45
Greedy decoding (n=500 per relation)	**61.20**	**69.80**	**80.00**	77.00	53.00	**89.60**	**85.60**	**92.20**	**89.40**	**77.53**
Human validation of gold ATOMIC	84.62	86.13	83.12	78.44	83.92	91.37	81.98	95.18	90.90	86.18

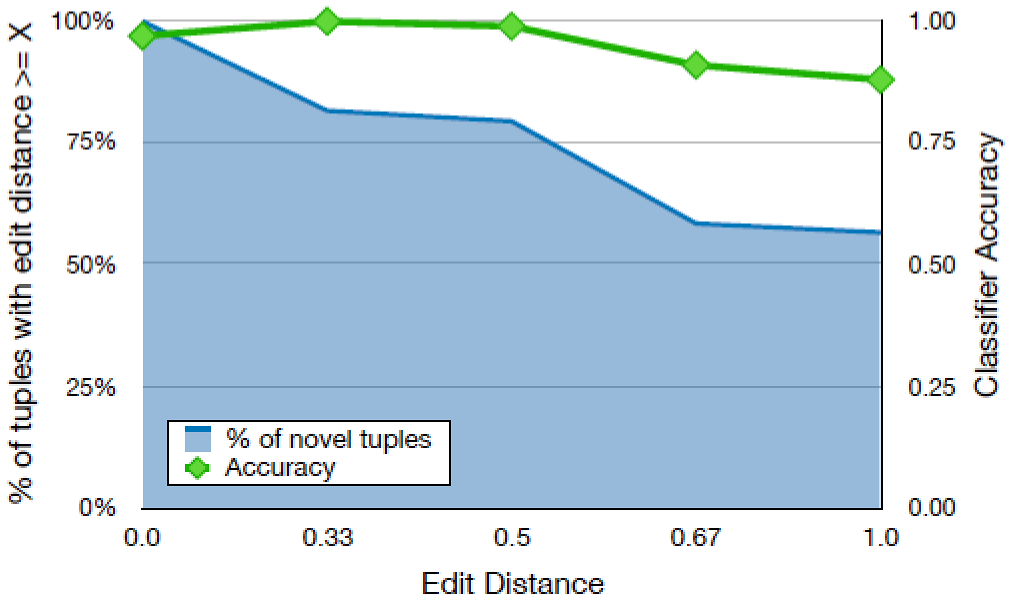
% train data	PPL	BLEU-2	N/T o	N/U o
1% train	23.81	5.08	7.24	49.36
10% train	13.74	12.72	**9.54**	**58.34**
50% train	11.82	13.97	9.32	50.37
FULL (- pretrain)	15.18	13.22	7.14	44.55
FULL train	**11.13**	**14.34**	9.51	50.05

Experiments

[ConceptNet]

Model	PPL	Score	N/T *sro*	N/T *o*	Human
LSTM - *s*	-	60.83	**86.25**	7.83	63.86
CKBG (Saito et al., 2018)	-	57.17	**86.25**	**8.67**	53.95
COMET (- pretrain)	8.05	89.25	36.17	6.00	83.49
COMET - RelTok	4.39	95.17	56.42	2.62	**92.11**
COMET	**4.32**	**95.25**	59.25	3.75	91.69

Table 6: ConceptNet generation Results



Conclusions

1. Introducing COMET for automatic construction of common-sense knowledge bases.

2. Adapting the weights of language models to learn to produce novel and diverse commonsense knowledge tuples.

3. Empirical results show that COMET can produces novel commonsense knowledges.

Q & A