# Commonsense Knowledge
# in
# Language Understanding & Generation

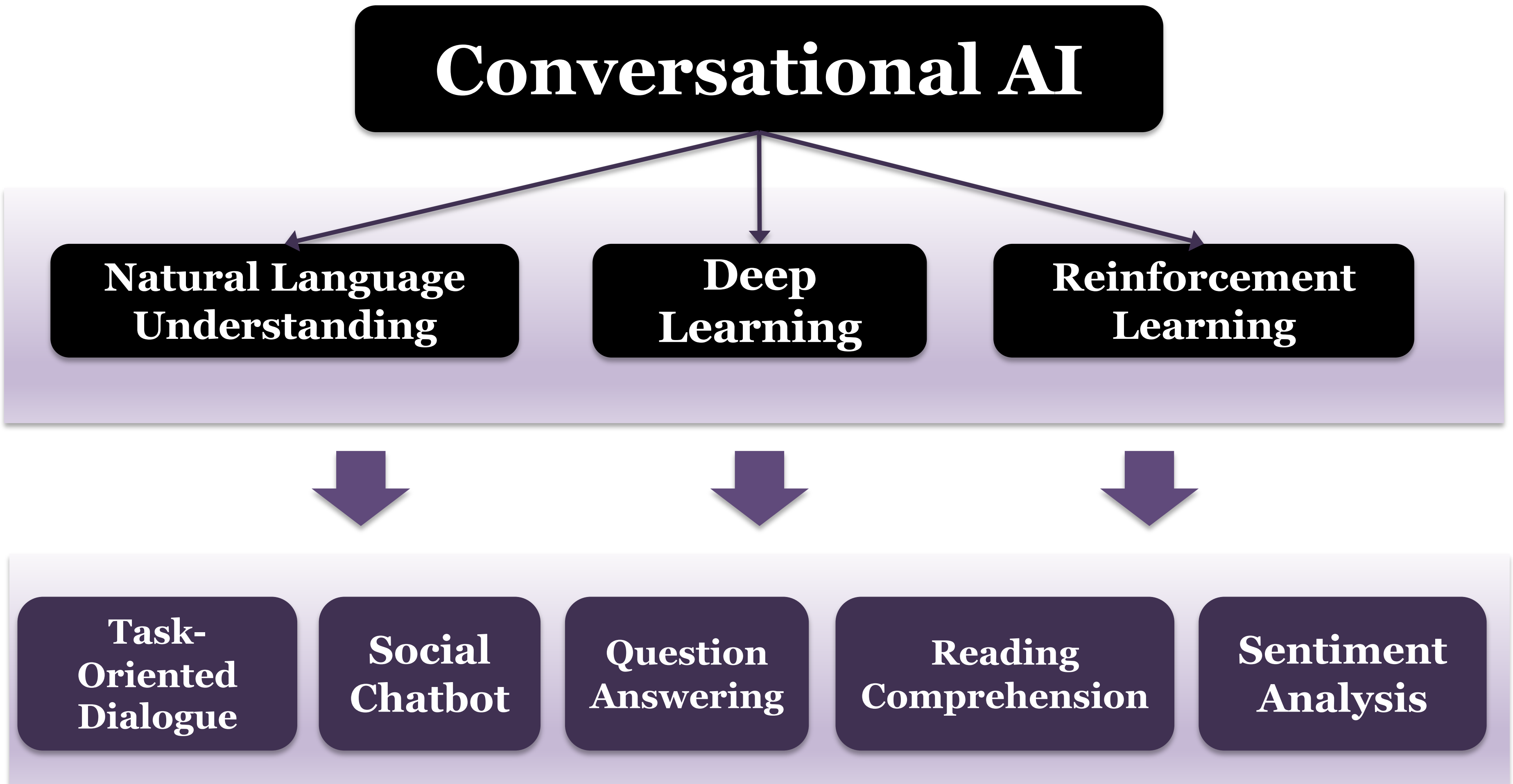

**Dr. Minlie Huang (黄民烈)**

**CS Department, Tsinghua University**

**aihuang@tsinghua.edu.cn**

**http://coai.cs.tsinghua.edu.cn/hml**

# Research Picture of the CoAI Group



# Outline

- Background: Knowledge
- Commonsense (CS) Extraction
- CS in Language Inference, Commonsense Reasoning
- CS in Machine Reading Comprehension
- CS in Language Generation (Story, Dialogue, etc.)

# Knowledge

- Knowledge is a familiarity, awareness, or understanding of someone or something, such as facts, information, descriptions, or skills, which is acquired through experience or education by perceiving, discovering, or learning.
- Knowledge can refer to a theoretical or practical understanding of a subject
- Plato(柏拉图): justified, true, believed
- Francis Bacon: Knowledge is Power

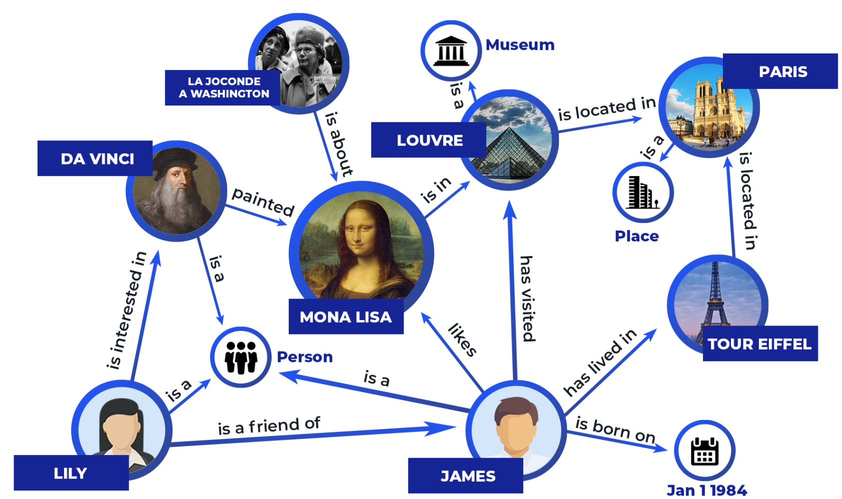
# Knowledge Types

- **World facts**
- **Commonsense knowledge**
- **Model knowledge** (prior distribution, knowledge distillation, knowledge transfer)

# Knowledge – world facts

- Knowledge graph by Google on May 16, 2012
- 70 billion facts (Oct. 2016); support search, Google Assistant, Google Home
- Knowledge triples (head entity, relation, tail entity)

(DA Vinci, painted, Mona Lisa)
(Mona Lisa, is in, Louvre)
(Louvre, locatedIn, Paris)

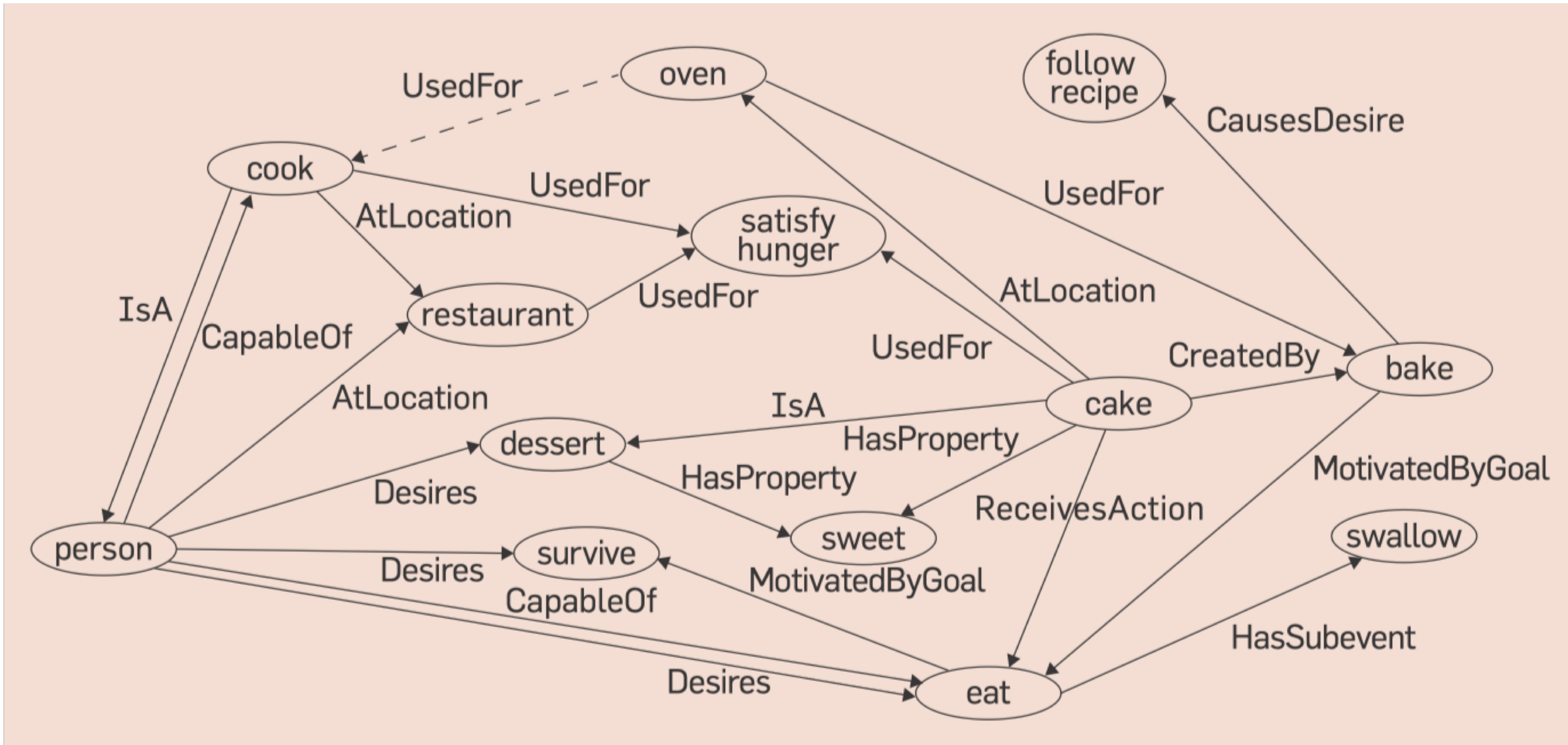


# Knowledge– commonsense

- **Commonsense knowledge** consists of facts about the everyday world, that all humans are expected to know. (Wikipedia)
  - Lemons are sour
  - Tree has leaves
  - Dog has four legs
- Commonsense knowledge bases
  - ConceptNet
  - Cyc
  - Open Mind Common Sense

# Knowledge– commonsense

- ConceptNet: 21 language-independent relations



# Knowledge– commonsense

- ConceptNet: 21 language-independent relations

•IsA
•UsedFor
•HasA
•CapableOf
•Desires
•CreatedBy ("cake" can be created by "baking")
•PartOf
•Causes
•LocatedNear
•AtLocation (Somewhere a "cook" can be at a "restaurant")
•DefinedAs
•SymbolOf (*X* represents *Y*)
•ReceivesAction ("cake" can be "eaten")
•HasPrerequisite (*X* can't do *Y* unless *A* does *B*)
•MotivatedByGoal (You would "bake" because you want to "eat")
•CausesDesire ("baking" makes you want to "follow recipe")
•MadeOf
•HasFirstSubevent (The first thing required when you're doing *X* is for entity *Y* to do *Z*)
•HasSubevent ("eat" has subevent "swallow")
•HasLastSubevent

# Knowledge– commonsense

- **Winograd Schema Challenge**: An alternative to Turing Test (**fooling** human judges vs. **testing** machine's intelligence)
- A **Winograd schema** is a pair of sentences that contain an ambiguity which requires world knowledge or reasoning to resolve it.
    - The city councilmen refused the demonstrators a permit because **they** [feared/advocated] violence.

# Knowledge – commonsense

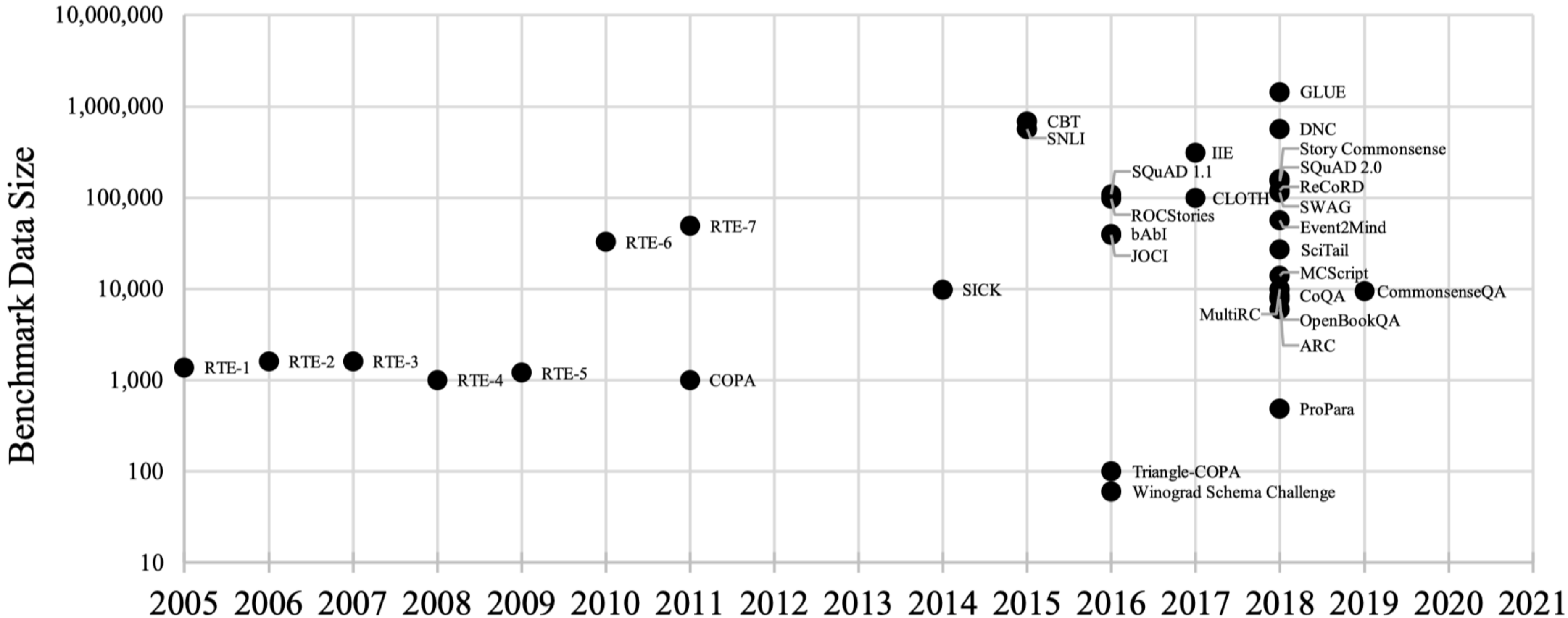
- "The Winograd Schema Challenge" Hector Levesque (2012):
  - Easily disambiguated by the human reader (ideally, so easily that the reader does not even notice that there is an ambiguity);
  - Not solvable by simple techniques such as *selectional restrictions*;
  - Google-proof; there is no obvious statistical test over text corpora that will reliably disambiguate these correctly.
- 150 schemas: https://cs.nyu.edu/faculty/davise/papers/WinogradSchemas/WSCollection.html
- Examples
  - The trophy would not fit in the brown suitcase because it was too **big**/***small***. What was too **big**/***small***?

Levesque et al. The Winograd Schema Challenge. AAAI 2012.

# Knowledge – commonsense

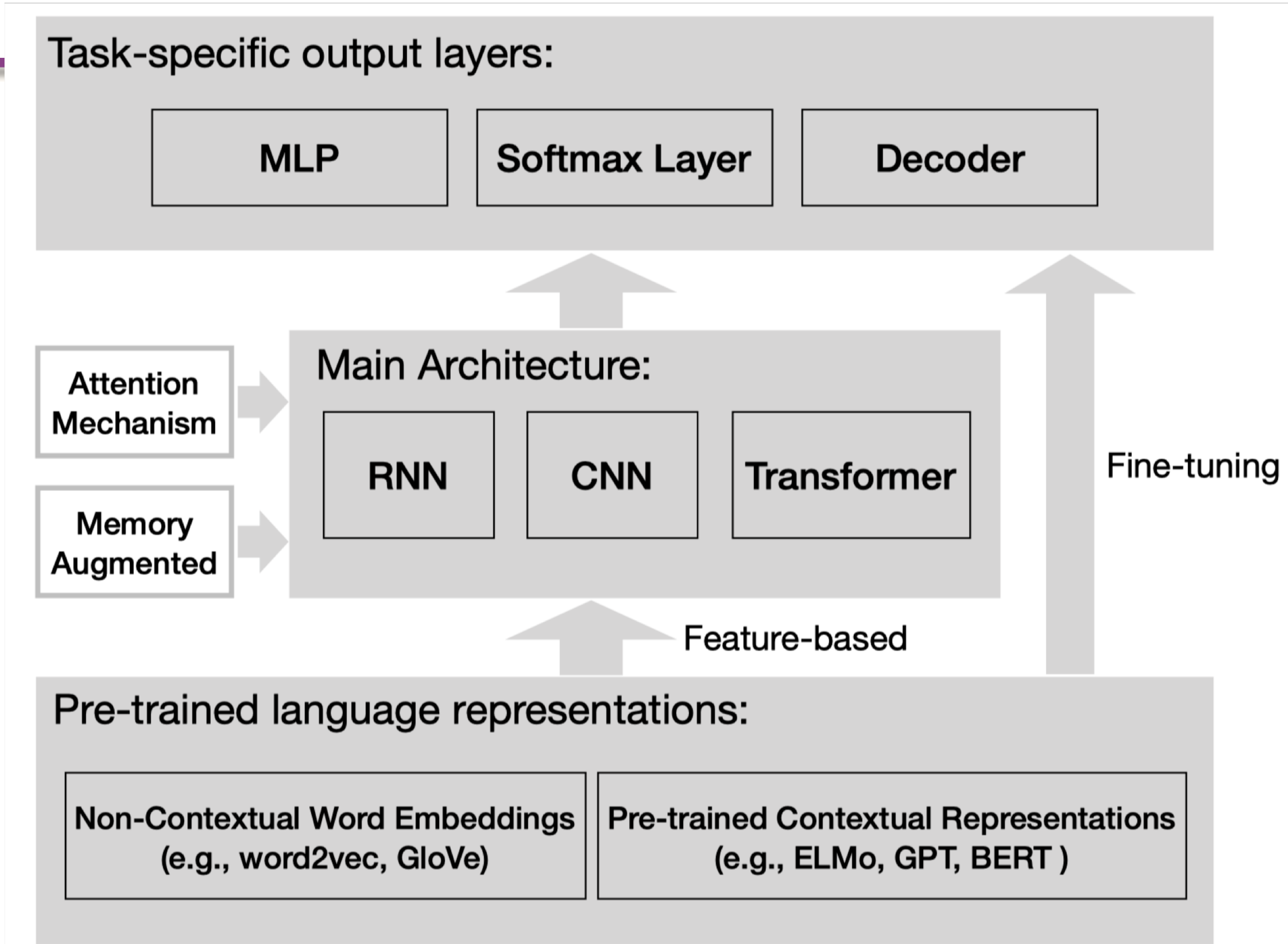
- A program has common sense if it automatically deduces for itself a sufficiently wide class of immediate consequences of anything it is told and what it already knows. – **McCarthy** (1959)

# Recent Advances




Stokes et al. 2019. Commonsense Reasoning for Natural Language Understanding: A Survey of Benchmarks, Resources, and Approaches

# Recent Advances




Stokes et al. 2019. Commonsense Reasoning for Natural Language Understanding: A Survey of Benchmarks, Resources, and Approaches

# Knowledge
## Extraction

# Commonsense Extraction

- We all know about it but we don't speak it out
- Resources
  - From crowd workers
  - From embeddings [1]
  - As knowledge base completion [2]
  - From raw data (text, image) [3]

① Yang et al. 2018. Extracting Commonsense Properties from Embeddings with Limited Human Guidance
② Li et al. 2018. Commonsense Knowledge Base Completion
③ Xu et al. 2018. Automatic Extraction of Commonsense LocatedNear Knowledge

# Commonsense Knowledge Base Completion

| relation | right term | conf. |
|---|---|---|
| MotivatedByGoal | relax | 3.3 |
| UsedFor | relaxation | 2.6 |
| MotivatedByGoal | your muscle be sore | 2.3 |
| HasPrerequisite | go to spa | 2.0 |
| Causes | get pruny skin | 1.6 |
| HasPrerequisite | change into swim suit | 1.6 |

Table 1: ConceptNet tuples with left term "soak in hotspring"; final column is confidence score.

$$\begin{aligned} \mathrm{loss}_{\mathrm{hinge}}(\tau) = \\ &\max\{0, \gamma - \mathrm{score}(\tau) + \mathrm{score}(\tau_{neg(t_1)})\} \\ &+ \max\{0, \gamma - \mathrm{score}(\tau) + \mathrm{score}(\tau_{neg(R)})\} \\ &+ \max\{0, \gamma - \mathrm{score}(\tau) + \mathrm{score}(\tau_{neg(t_2)})\} \end{aligned}$$

Bilinear model

Deep neural models

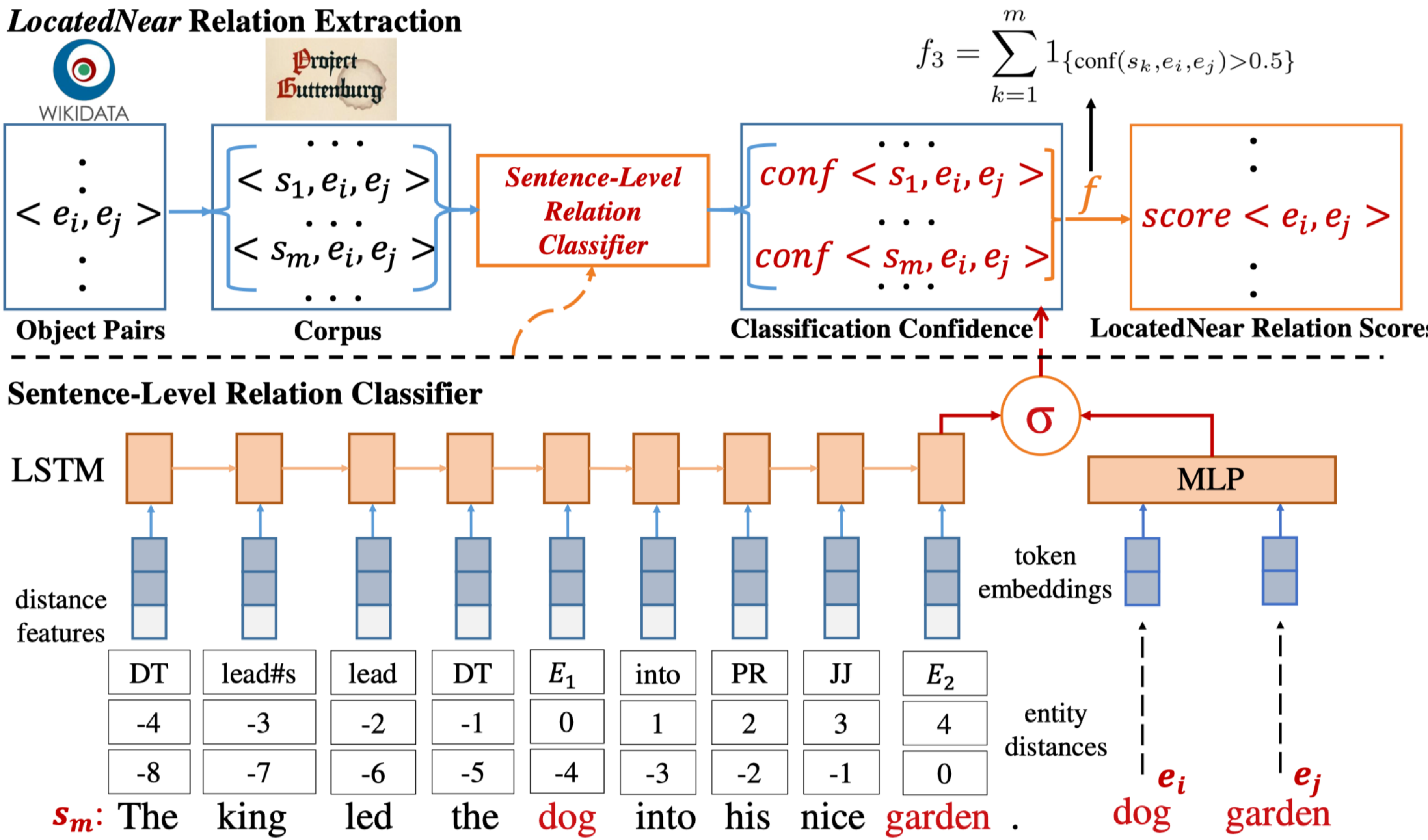
# LocatedNear Knowledge

- Extract LocatedNear relation from text
- Why
  - Object detection
  - RC for spatial facts and physical scenes
  - ConceptNet 5.5 has only 49 triples of this relation

Figure 1: LOCATEDNEAR facts assist the detection of vague objects: if a set of knife, fork and plate is on the table, one may believe there is a glass beside based on the commonsense, even though these objects are hardly visible due to low light.

Xu et al. 2018. Automatic Extraction of Commonsense LocatedNear Knowledge

# LocatedNear Knowledge

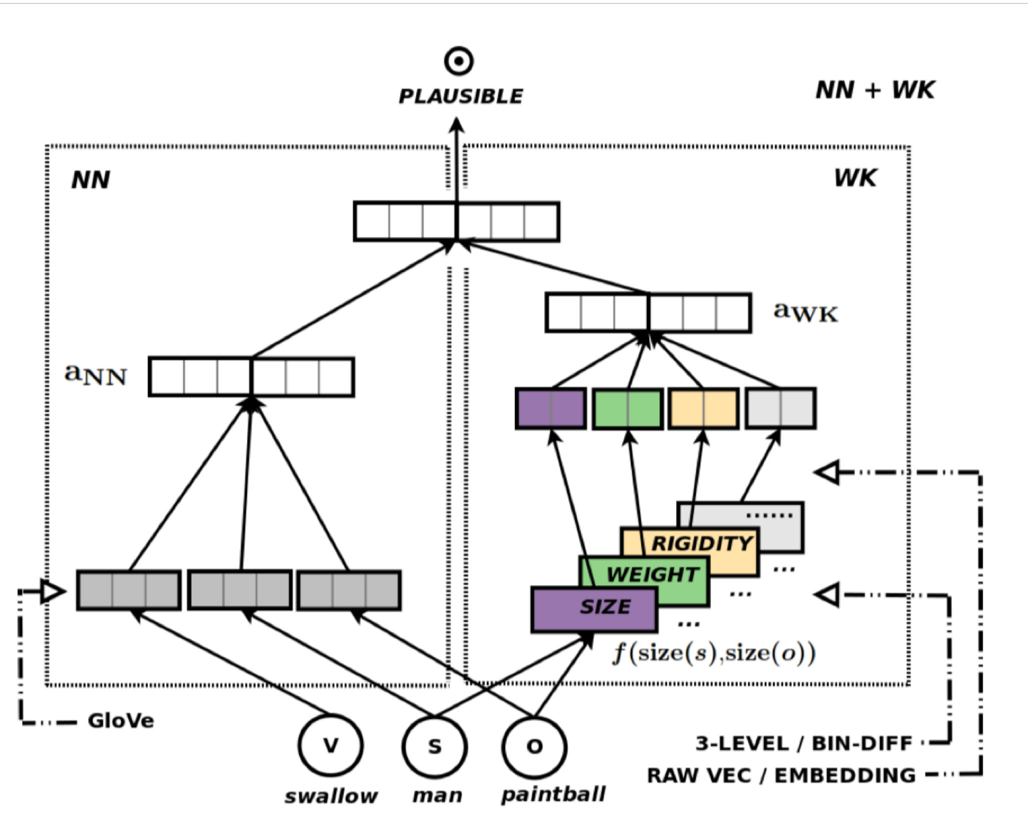


# Injecting World Knowledge

| *man-swallow-** | PREFERRED? | PLAUSIBLE? |
| --- | --- | --- |
| *-candy* | ✓ | ✓ |
| *-paintball* | ✗ | ✓ |
| *-desk* | ✗ | ✗ |

Wang et al. NAACL 2018. Modeling Semantic Plausibility by Injecting World Knowledge

# Injecting World Knowledge




- SENTIENCE: *rock, tree, ant, cat, chimp, man*.
- MASS-COUNT: *milk, sand, pebbles, car*.
- PHASE: *smoke, milk, wood*.
- SIZE: *watch, book, cat, person, jeep, stadium*.
- WEIGHT: *watch, book, dumbbell, man, jeep, stadium*.
- RIGIDITY: *water, skin, leather, wood, metal*.

Wang et al. NAACL 2018. Modeling Semantic Plausibility by Injecting World Knowledge

# Ordinal Commonsense Inference

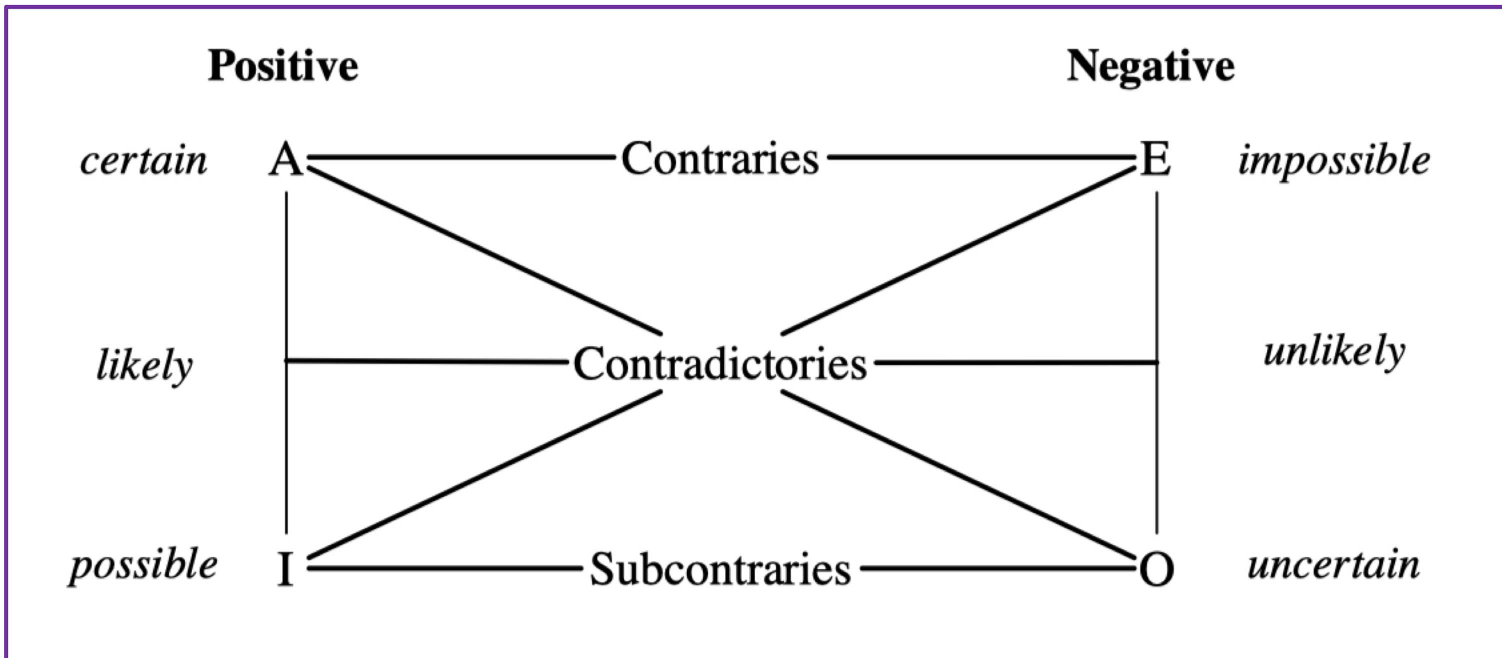
## Plausible but not entailed

T: A person flips a coin.

H: That flip comes up heads.

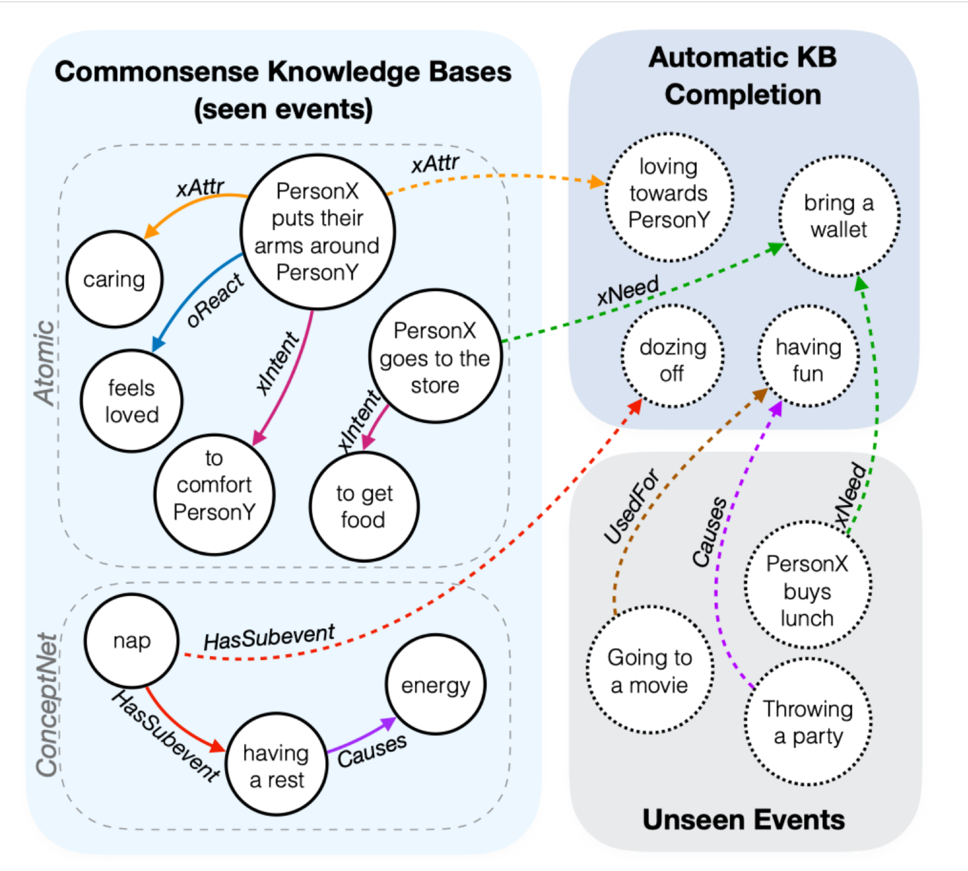
T: An animal eats food.

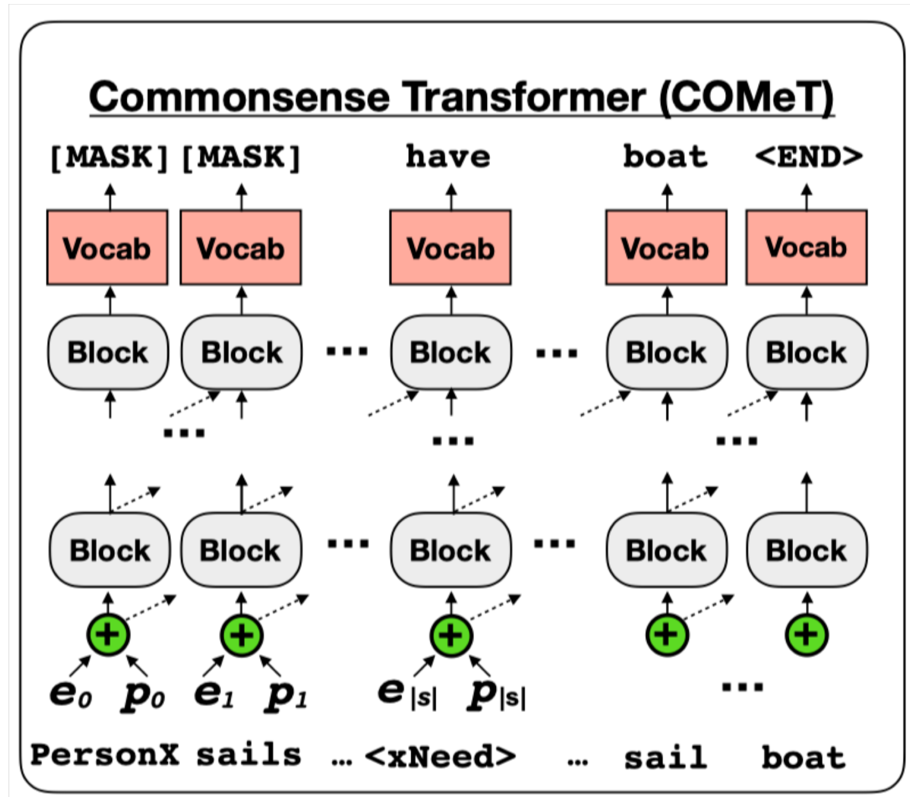
H: A person eats food.




Zhang et al. TACL 2017. Ordinal Common-sense Inference

# Comet: Commonsense Transformers




Bosselut et al. 2019. COMET : Commonsense Transformers for Automatic Knowledge Graph Construction

# Knowledge
# in
# Language Inference

# Inference with CS Knowledge

- Natural language inference
  - **Premise**: A lady standing in a wheat field.
  - **Hypothesis**: A person standing in a corn field.
- Commonsense reasoning
  - The trophy doesn't fit in the suitcase because it is too big. What is too big?
    **Answer 0**: the trophy
    **Answer 1**: the suitcase

# Natural Language Inference

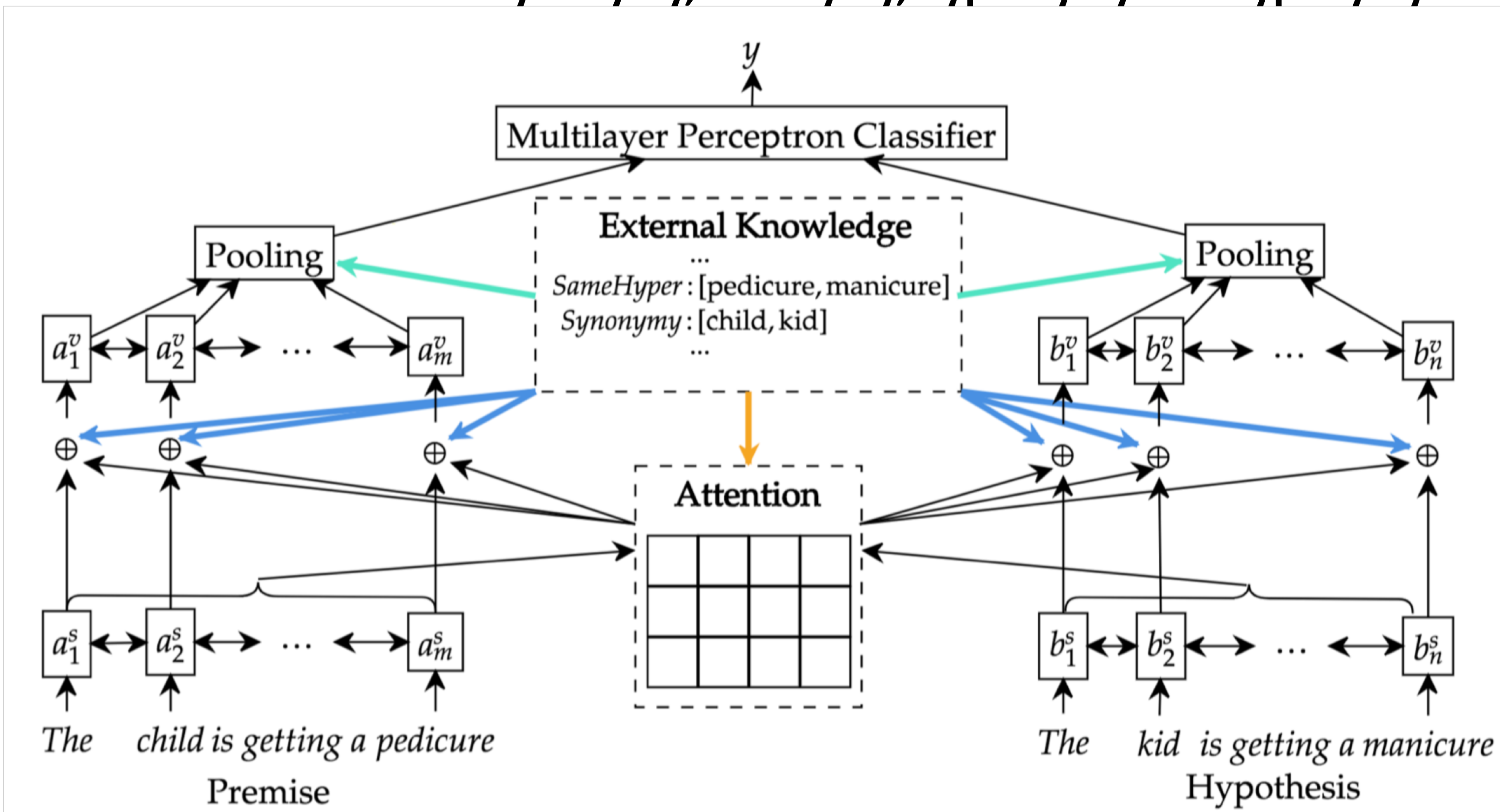
| | | |
|---|---|---|
| A man inspects the uniform of a figure in some East Asian country. | **contradiction**<br>C C C C C | The man is sleeping |
| An older and younger man smiling. | **neutral**<br>N N E N N | Two men are smiling and laughing at the cats playing on the floor. |
| A black race car starts up in front of a crowd of people. | **contradiction**<br>C C C C C | A man is driving down a lonely road. |
| A soccer game with multiple males playing. | **entailment**<br>E E E E E | Some men are playing a sport. |
| A smiling costumed woman is holding an umbrella. | **neutral**<br>N N E C N | A happy woman in a fairy costume holds an umbrella. |


Bowman S R, Angeli G, Potts C, et al. A large annotated corpus for learning natural language inference[J]. EMNLP 2015.

Williams A, Nangia N, Bowman S R. A broad-coverage challenge corpus for sentence understanding through inference[J]. NAACL 2017.

# Natural Language Inference

**synonymy, antonymy, hypernymy and hyponymy**





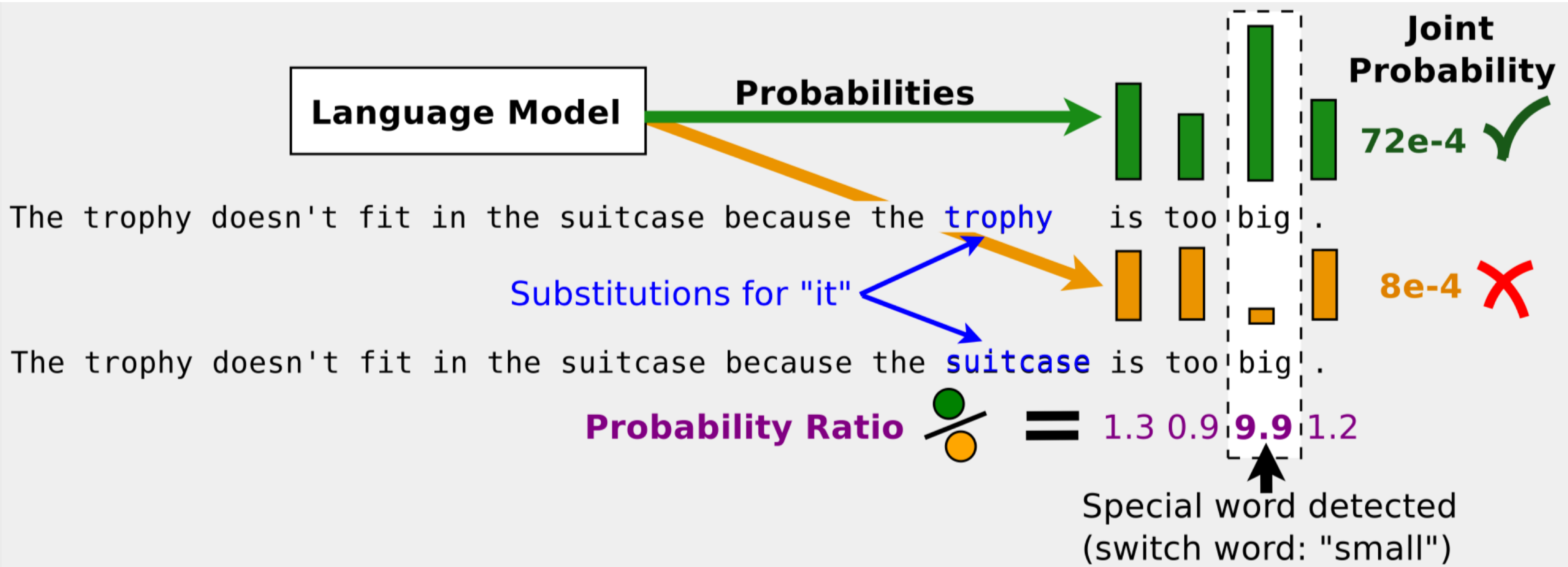

Chen et al. ACL 2018. Neural Natural Language Inference Models Enhanced with External Knowledge.

# Winograd Schema Challenge

- Winograd schema
  - **Two parties mentioned**
  - **A pronoun or possessive adjective**
  - The question involves determining the referent of the pronoun or possessive adjective
  - If alternate word is replaced, answer changes

The city councilmen refused the demonstrators a permit because **they** [feared/advocated] violence.
Who feared/advocated violence?

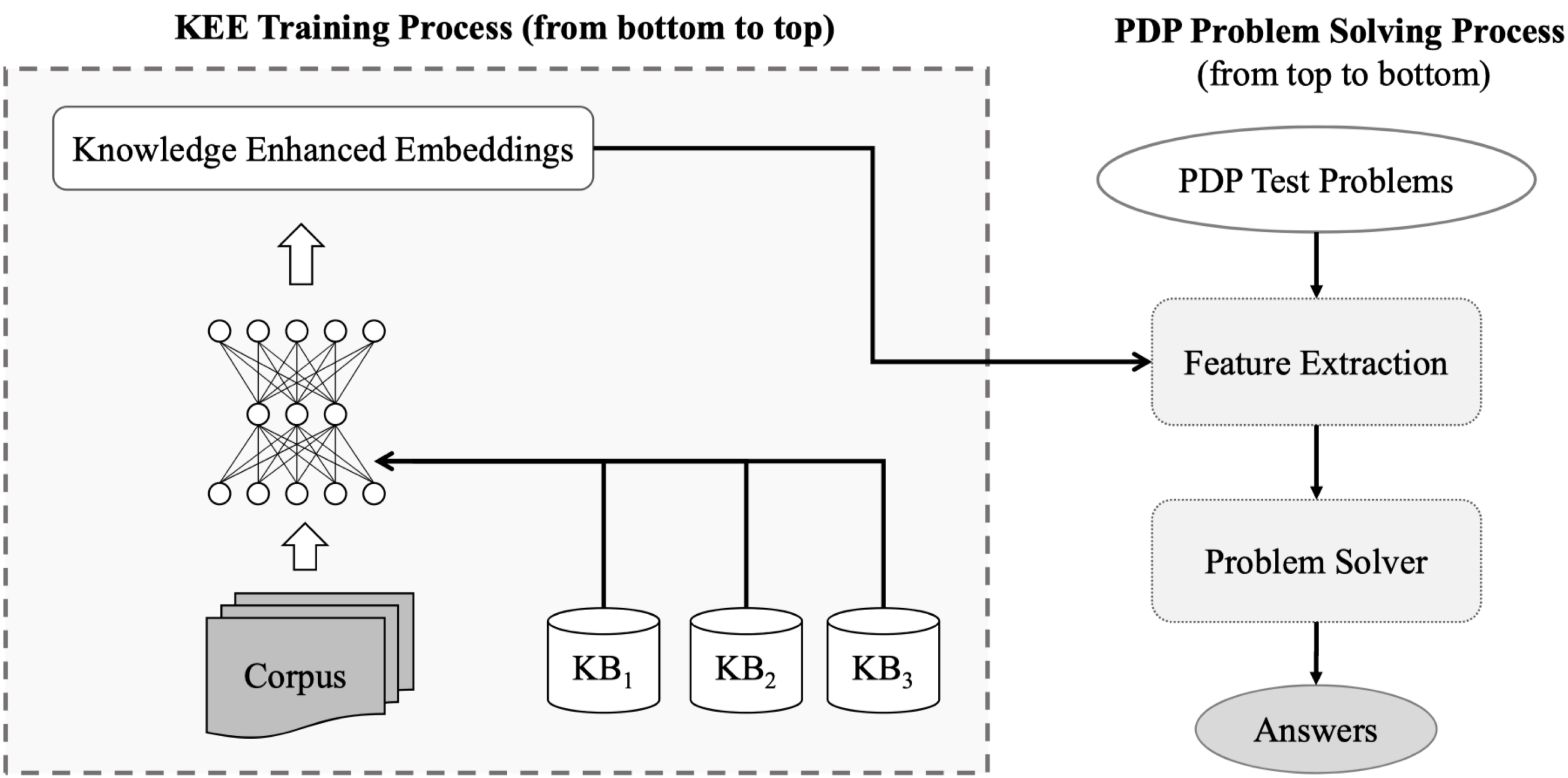
# Commonsense Reasoning with LM




PDP-60: language model ensembles 70%
WSC-273: language model ensembles 63%

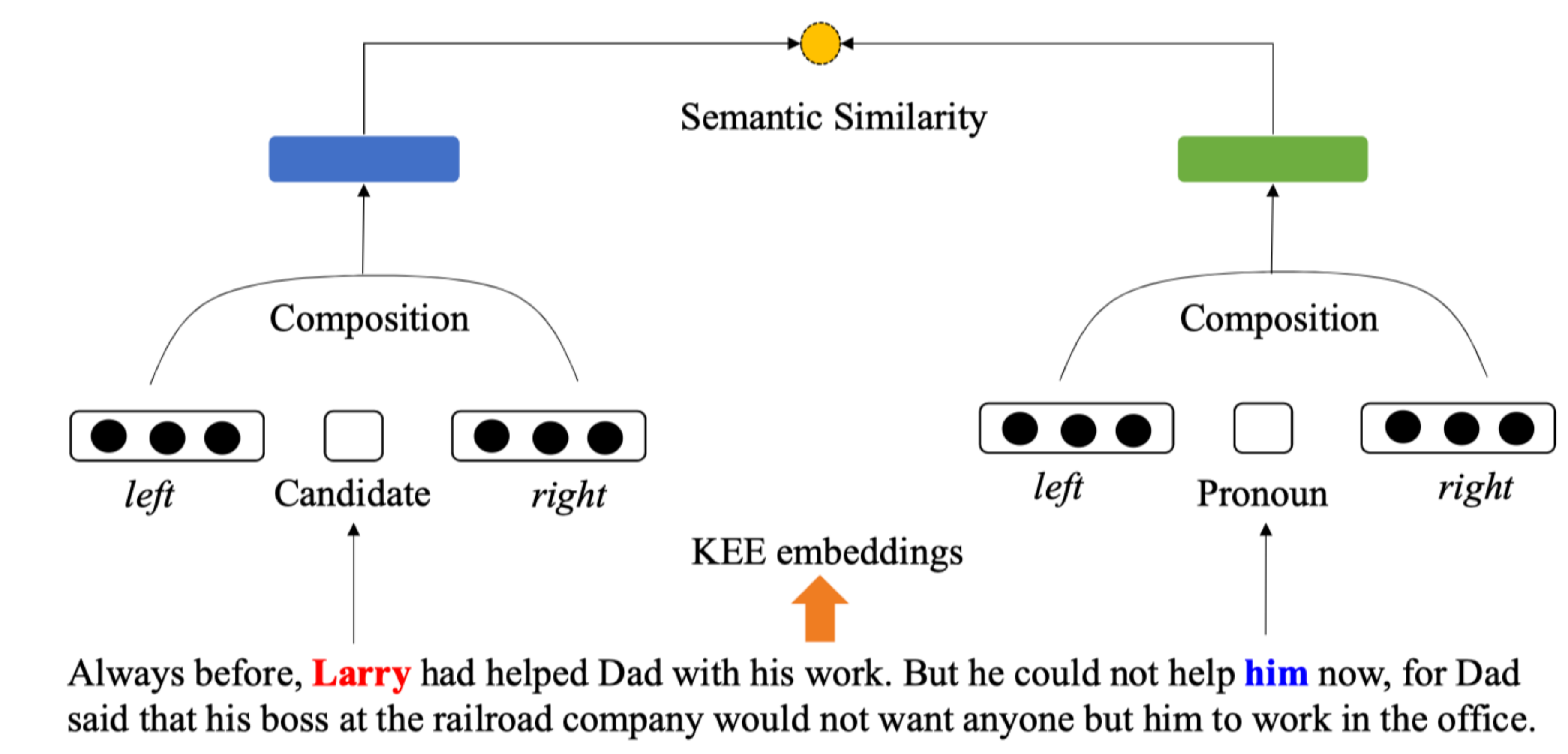
Trinh and Le. A Simple Method for Commonsense Reasoning

# Knowledge Enhanced Embeddings



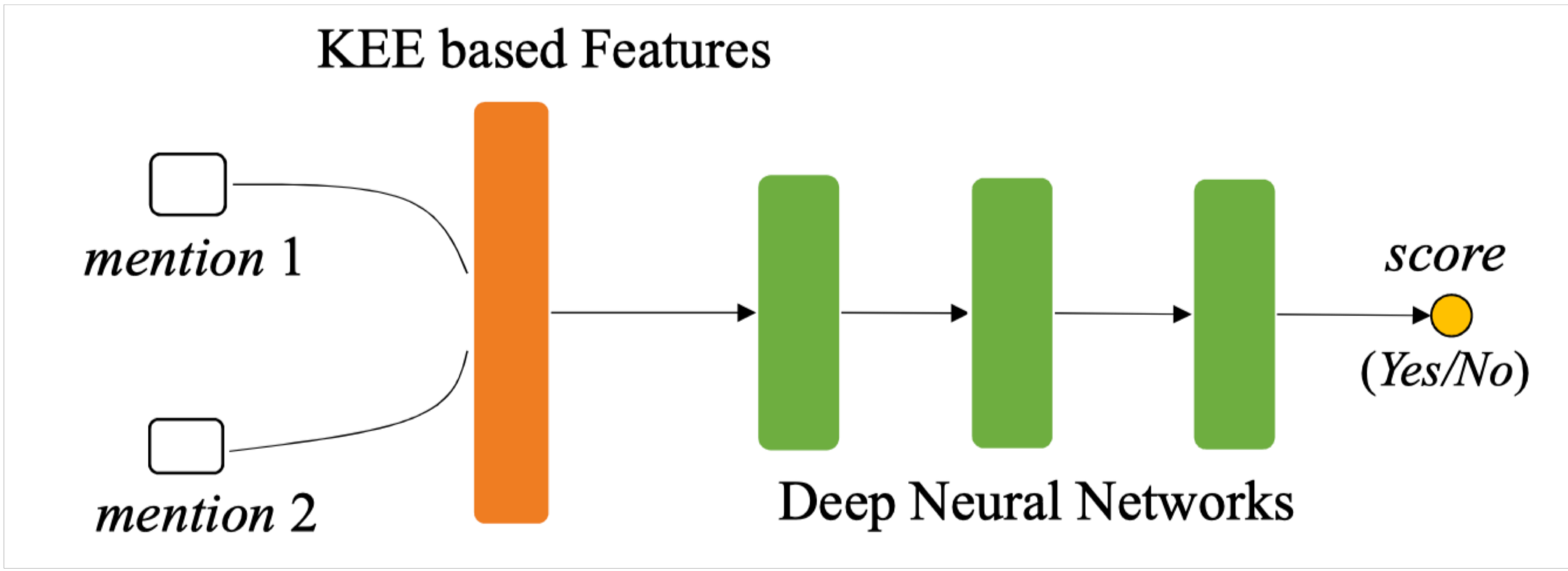

Liu et al. 2016. Commonsense Knowledge Enhanced Embeddings for Solving Pronoun Disambiguation Problems in Winograd Schema Challenge

# Knowledge Enhanced Embeddings



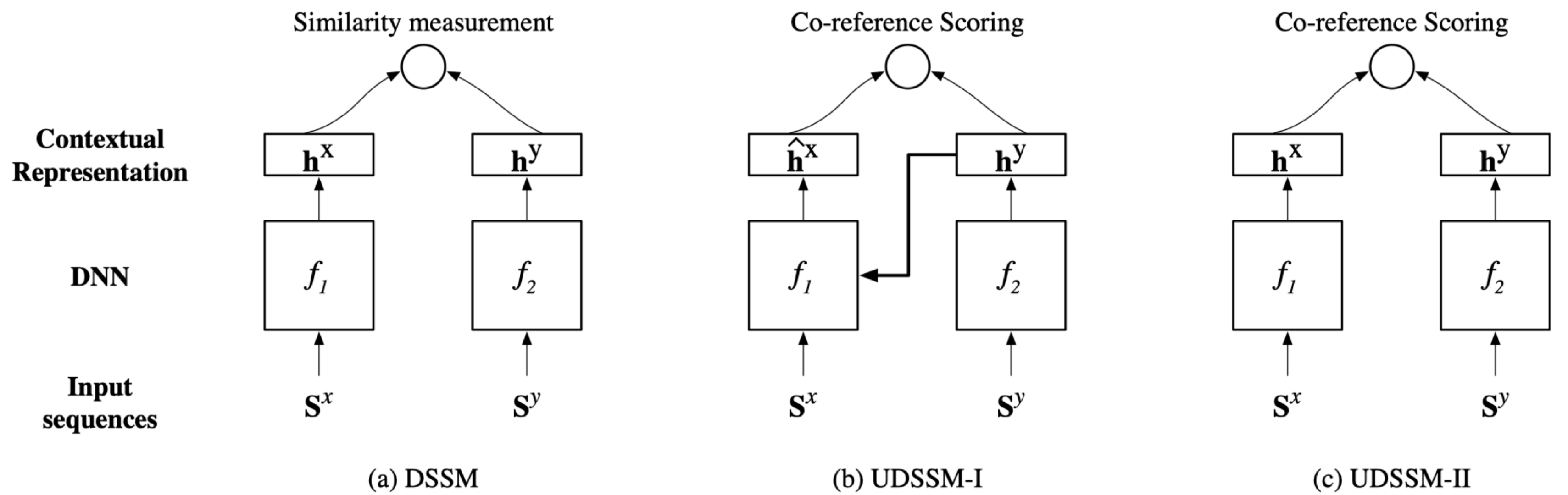

Liu et al. 2016. Commonsense Knowledge Enhanced Embeddings for Solving Pronoun Disambiguation Problems in Winograd Schema Challenge

# Knowledge Enhanced Embeddings




Liu et al. 2016. Commonsense Knowledge Enhanced Embeddings for Solving Pronoun Disambiguation Problems in Winograd Schema Challenge

# Unsupervised Deep Structured Semantic Models for Commonsense Reasoning




| | | |
|---|---|---|
| ELMo | 56.7% | 51.5% |
| Google Language Model (Trinh and Le, 2018) | 60.0% | 56.4% |
| **UDSSM-I** | 75.0% | 54.5% |
| **UDSSM-II** | **75.0%** | **59.2%** |
| Google Language Model (ensemble) | 70.0% | 61.5% |
| UDSSM-I (ensemble) | 76.7% | 57.1% |
| UDSSM-II (ensemble) | **78.3%** | **62.4%** |

# Knowledge
# in
# Reading Comprehension

# An Evaluation of Commonsense Causal Reasoning (2011)

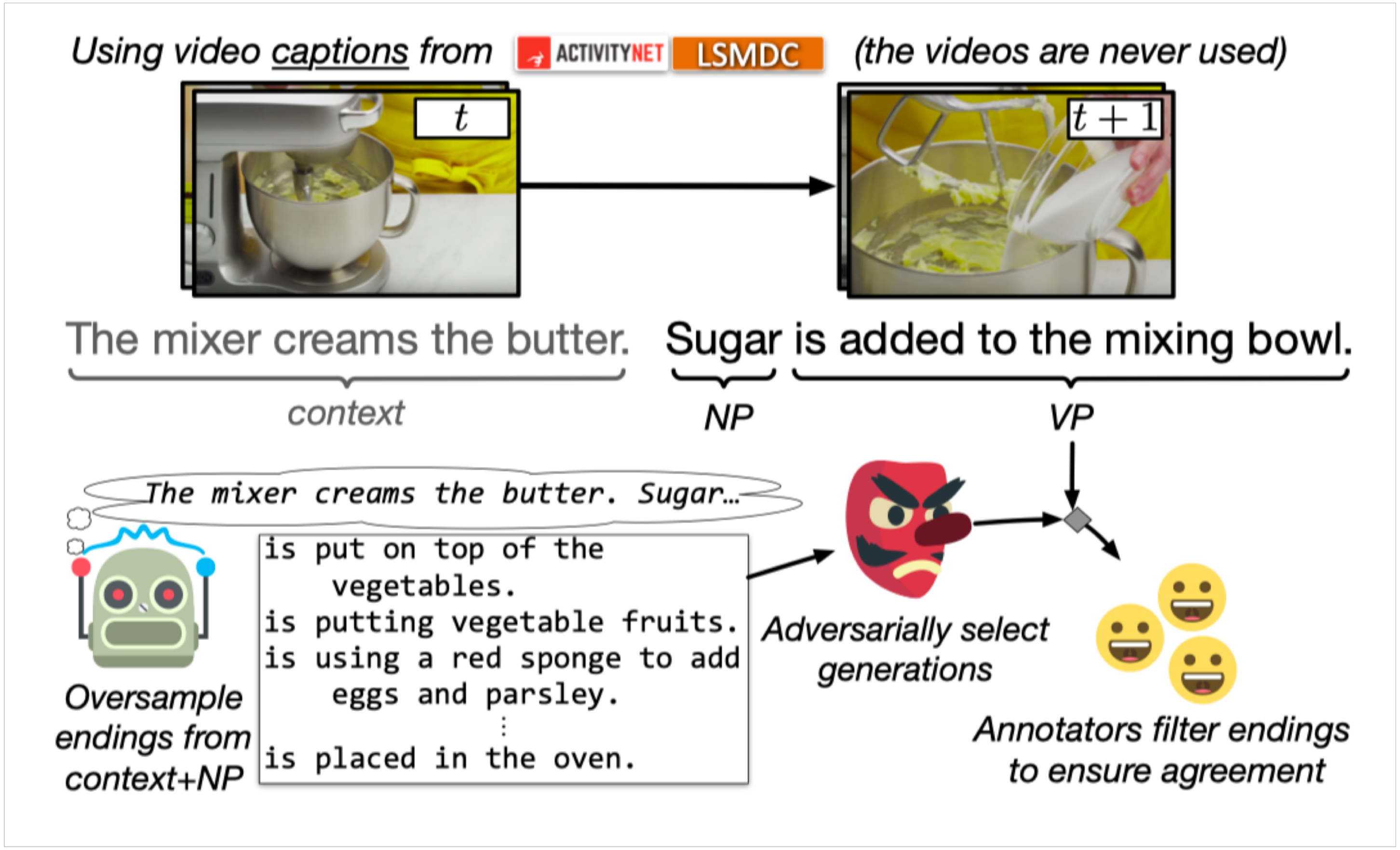
- Choice Of Plausible Alternatives (**COPA**): measuring the ability of resolving commonsense causality

(forward causal reasoning)
Premise: The man lost his balance on the ladder. *What happened as a result?*
Alternative 1: He fell off the ladder.
Alternative 2: He climbed up the ladder.

(backwards causal reasoning)
Premise: The man fell unconscious. *What was the cause of this?*
Alternative 1: The assailant struck the man in the head.
Alternative 2: The assailant took the man's wallet.


Roemmele et al. 2011. Choice of Plausible Alternatives: An Evaluation of Commonsense Causal Reasoning. AAAI 2011 spring symposium

# A Large-Scale Adversarial Dataset for Grounded Commonsense Inference




Zeller et al. EMNLP 2018. Swag: A Large-Scale Adversarial Dataset for Grounded Commonsense Inference.

# A Large-Scale Adversarial Dataset for Grounded Commonsense Inference

On stage, a woman takes a seat at the piano. She

a) sits on a bench as her sister plays with the doll.
b) smiles with someone as the music plays.
c) is in the crowd, watching the dancers.
**d) nervously sets her fingers on the keys.**

A girl is going across a set of monkey bars. She

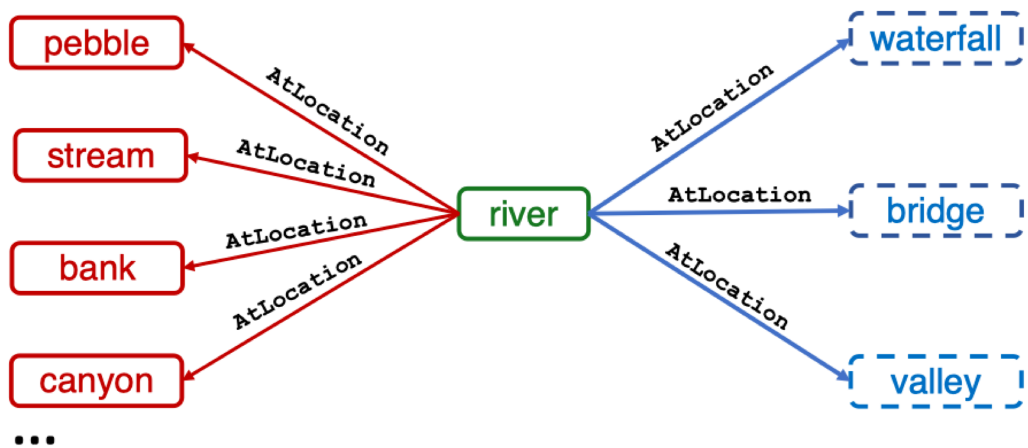
a) jumps up across the monkey bars.
b) struggles onto the monkey bars to grab her head.
**c) gets to the end and stands on a wooden plank.**
d) jumps up and does a back flip.

The woman is now blow drying the dog. The dog

**a) is placed in the kennel next to a woman's feet.**
b) washes her face with the shampoo.
c) walks into frame and walks towards the dog.
d) tried to cut her face, so she is trying to do something very close to her face.

# Commonsense QA

a) Sample ConceptNet for specific subgraphs




b) Crowd source corresponding natural language questions and two additional distractors

*Where on a **river** can you hold a cup upright to catch water on a sunny day?*

✓ waterfall, ✗ bridge, ✗ valley, ✗ *pebble*, ✗ **mountain**

*Where can I stand on a **river** to see water falling without getting wet?*

✗ waterfall, ✓ bridge, ✗ valley, ✗ *stream*, ✗ **bottom**

*I'm crossing the **river**, my feet are wet but my body is dry, where am I?*

✗ waterfall, ✗ bridge, ✓ valley, ✗ *bank*, ✗ **island**

Talmor et al. 2018. COMMONSENSE QA: A Question Answering Challenge Targeting Commonsense Knowledge

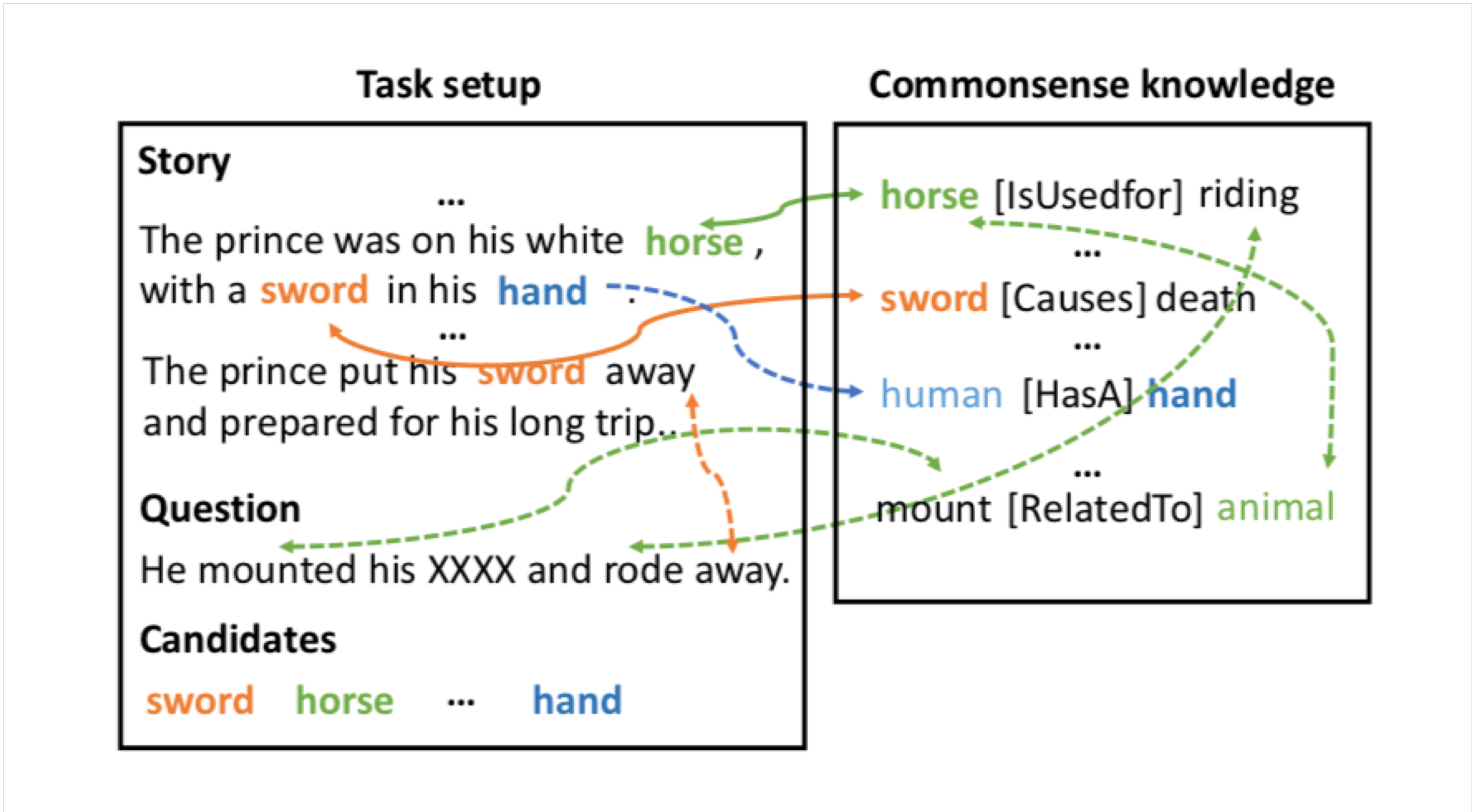
# Commonsense QA

| | Random split | | Question concept split | |
|---|---|---|---|---|
| Model | Accuracy | SANITY | Accuracy | SANITY |
| VECSIM+NUMBERBATCH | 29.1 | 54.0 | 30.3 | 54.9 |
| LM1B-REP | 26.1 | 39.6 | 26.0 | 39.1 |
| LM1B-CONCAT | 25.3 | 37.4 | 25.3 | 35.2 |
| VECSIM+GLOVE | 22.3 | 26.8 | 20.8 | 27.1 |
| BERT-LARGE | **55.9** | **92.3** | **63.6** | **93.2** |
| GPT | 45.5 | 87.2 | 55.5 | 88.9 |
| ESIM+ELMO | 34.1 | 76.9 | 37.9 | 77.8 |
| ESIM+GLOVE | 32.8 | 79.1 | 40.4 | 78.2 |
| QABILINEAR+GLOVE | 31.5 | 74.8 | 34.2 | 71.8 |
| ESIM+NUMBERBATCH | 30.1 | 74.6 | 31.2 | 75.1 |
| QABILINEAR+NUMBERBATCH | 28.8 | 73.3 | 32.0 | 71.6 |
| QACOMPARE+GLOVE | 25.7 | 69.2 | 34.1 | 71.3 |
| QACOMPARE+NUMBERBATCH | 20.4 | 60.6 | 25.2 | 66.8 |
| BIDAF++ | 32.0 | 71.0 | 38.4 | 72.0 |
| HUMAN | **88.9** | | | |



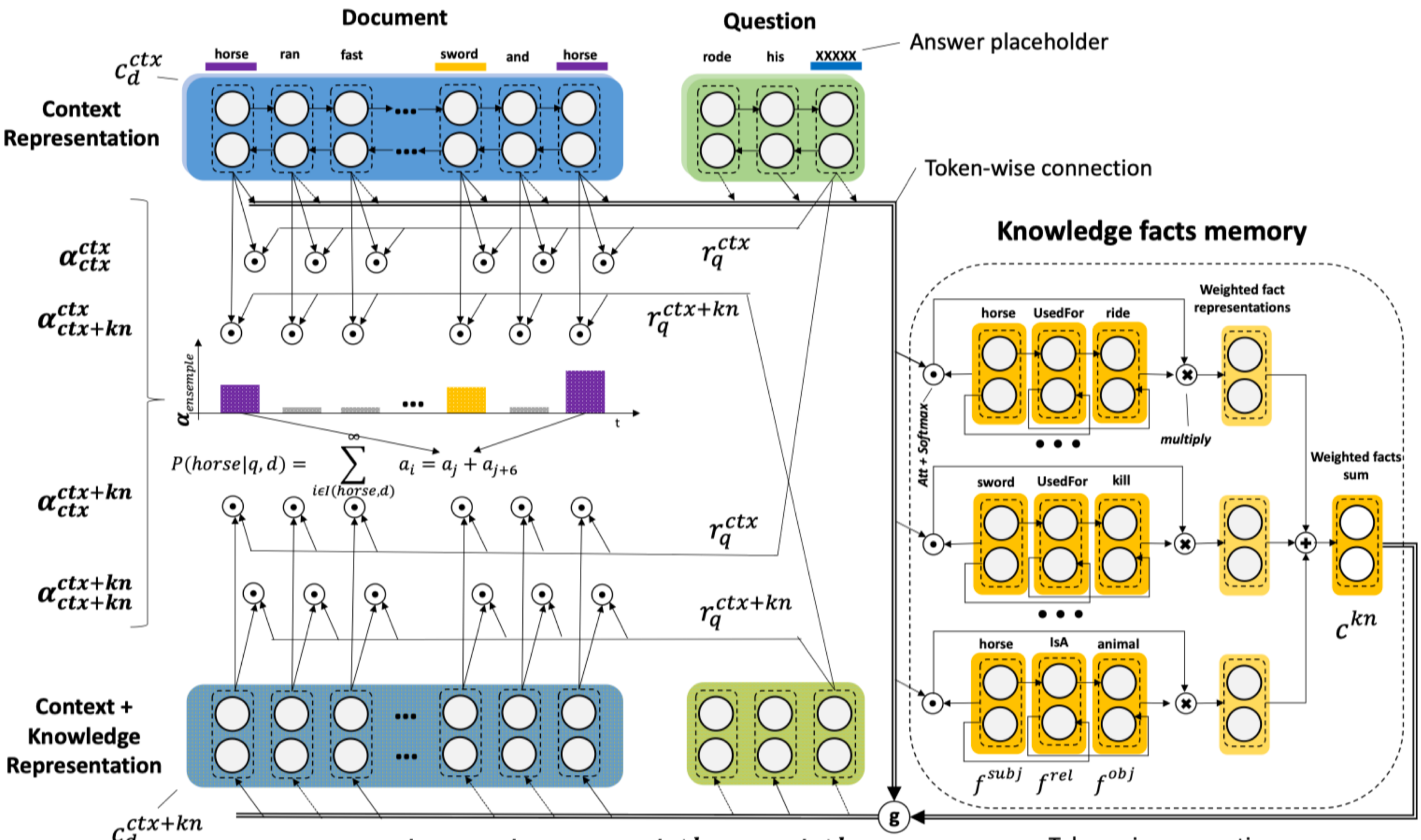
Talmor et al. 2018. COMMONSENSE QA: A Question Answering Challenge Targeting Commonsense Knowledge

# Knowledgeable Reader (Cloze Test)



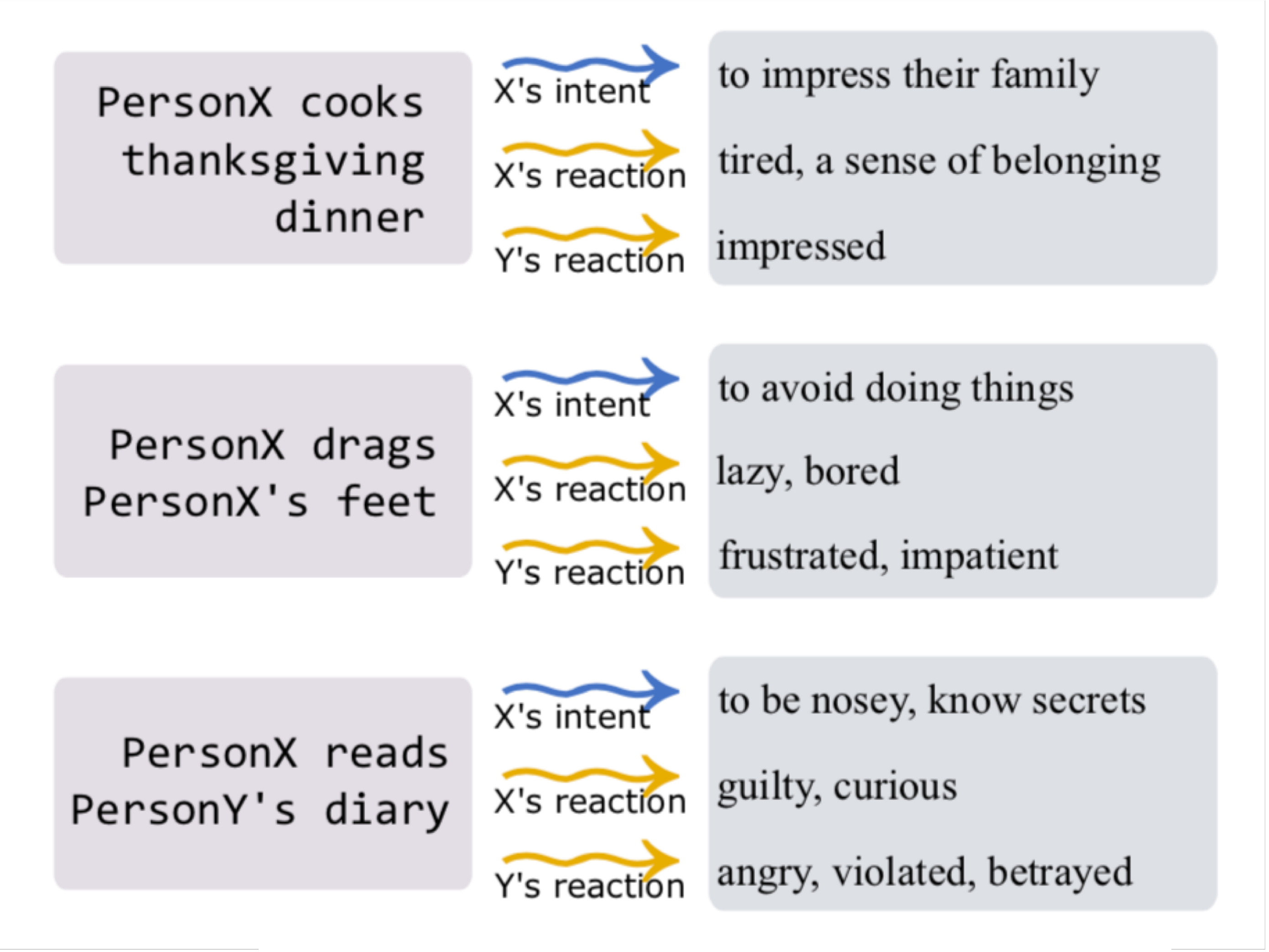

Mihaylov and Frank. 2018. Knowledgeable Reader: Enhancing Cloze-Style Reading Comprehension with External Commonsense Knowledge

# Knowledgeable Reader




$\alpha_{ensemble} = W_1\alpha_{ctx}^{ctx} + W_2\alpha_{ctx+kn}^{ctx} + W_3\alpha_{ctx}^{ctx+kn} + W_4\alpha_{ctx+kn}^{ctx+kn}$

# Event2Mind




Rashkin et al. 2018. Event2Mind: Commonsense Inference on Events, Intents, and Reactions

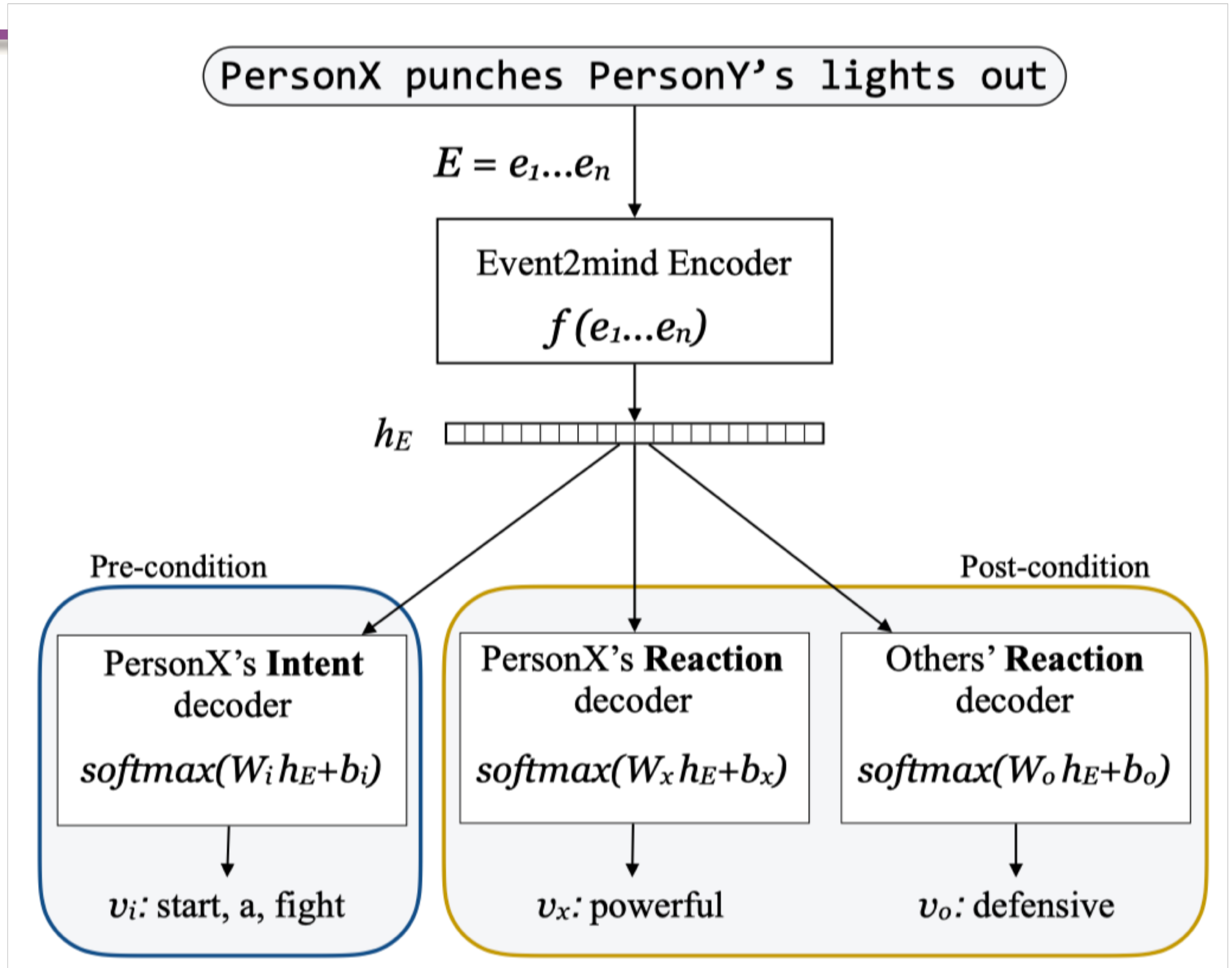
# Event2Mind

| PersonX's Intent | Event Phrase | PersonX's Reaction | Others' Reactions |
|---|---|---|---|
| to express anger<br>to vent their frustration<br>to get PersonY's full attention | **PersonX starts to yell at PersonY** | mad<br>frustrated<br>annoyed | shocked<br>humiliated<br>mad at PersonX |
| to communicate something without being rude<br>to let the other person think for themselves<br>to be subtle | **PersonX drops a hint** | sly<br>secretive<br>frustrated | oblivious<br>surprised<br>grateful |
| to catch the criminal<br>to be civilized<br>justice | **PersonX reports ___ to the police** | anxious<br>worried<br>nervous | sad<br>angry<br>regret |
| to wake up<br>to feel more energized | **PersonX drinks a cup of coffee** | alert<br>awake<br>refreshed | NONE |



Rashkin et al. 2018. Event2Mind: Commonsense Inference on Events, Intents, and Reactions

# Event2Mind

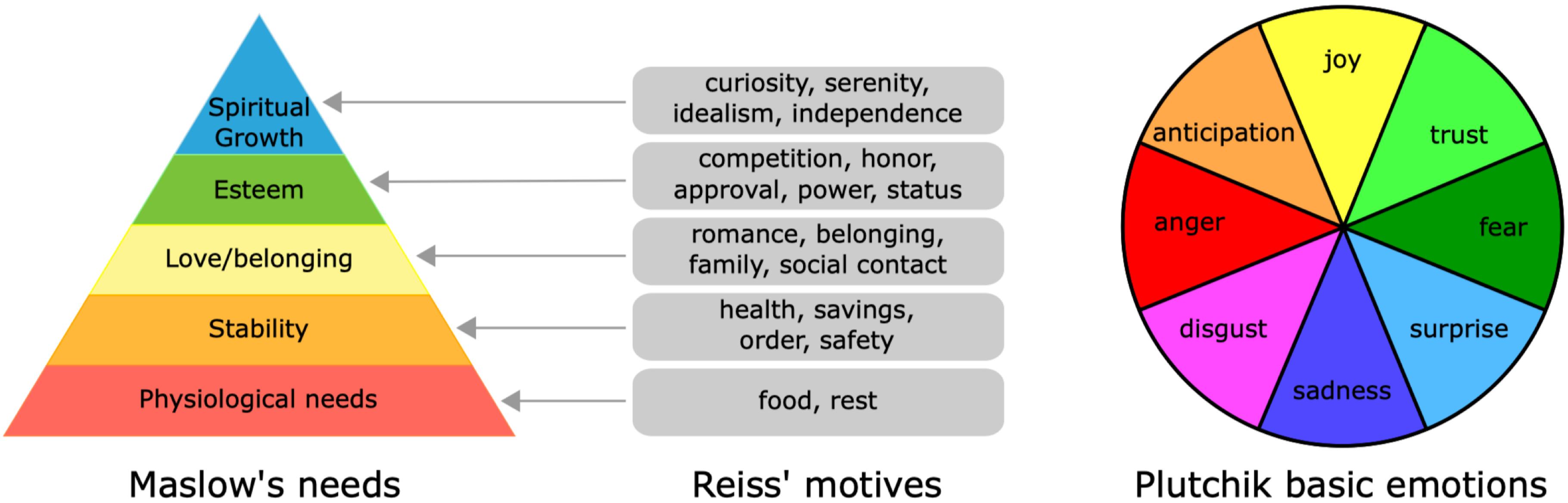





Rashkin et al. 2018. Event2Mind: Commonsense Inference on Events, Intents, and Reactions

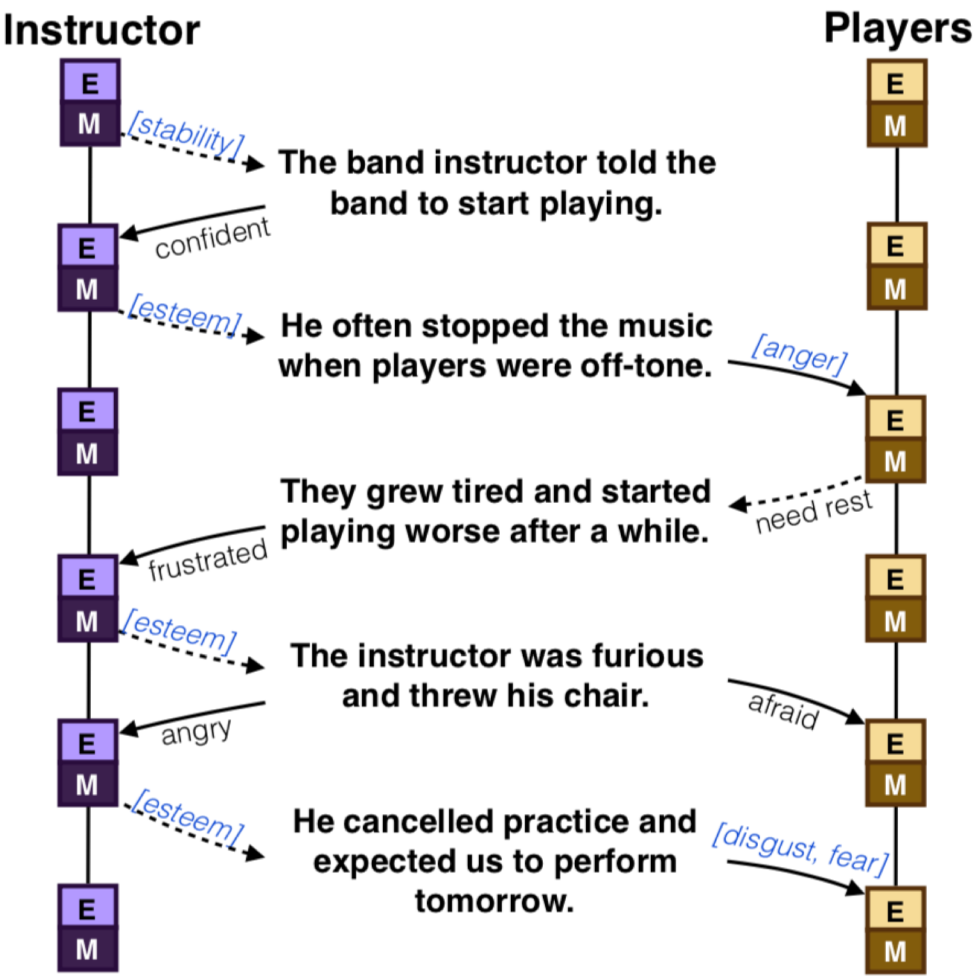
# Modeling Naive Psychology in commonsense stories

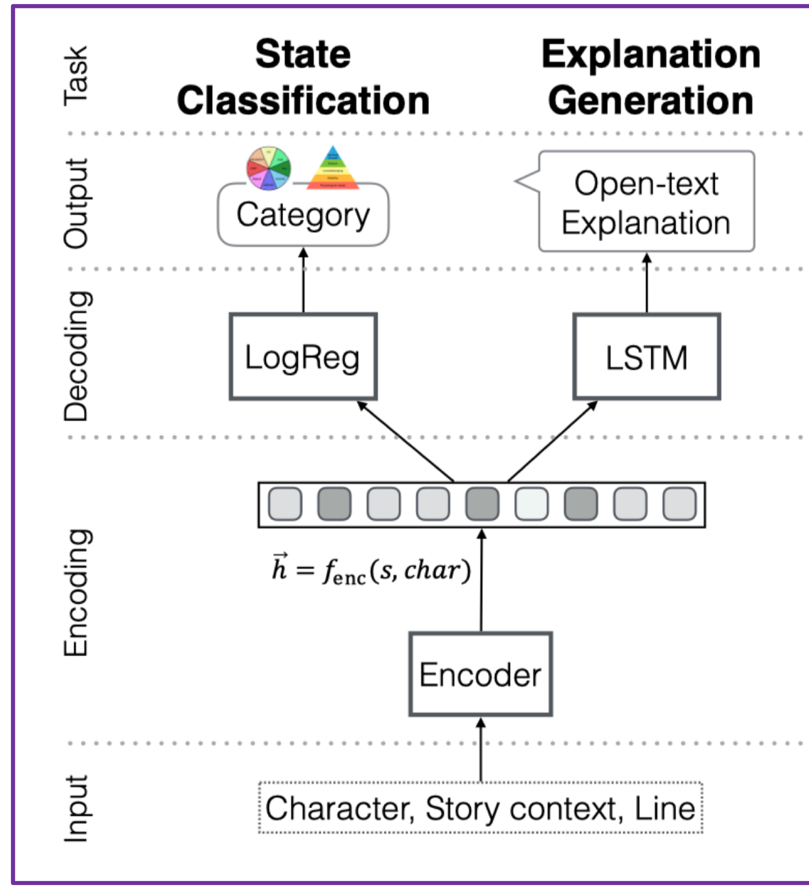
- Mental state: motivation and emotional reaction




Rashkin et al. 2018. Modeling Naive Psychology of Characters in Simple Commonsense Stories

# Modeling Naive Psychology in commonsense stories




Rashkin et al. 2018. Modeling Naive Psychology of Characters in Simple Commonsense Stories

# Cloze-style Machine Commonsense Reading Comprehension

**Passage**

(CNN) -- A lawsuit has been filed claiming that the iconic Led Zeppelin song "Stairway to Heaven" was far from original. The suit, filed on May 31 in the United States District Court Eastern District of Pennsylvania, was brought by the estate of the late musician Randy California against the surviving members of Led Zeppelin and their record label. The copyright infringement case alleges that the Zeppelin song was taken from the single "Taurus" by the 1960s band Spirit, for whom California served as lead guitarist. "Late in 1968, a then new band named Led Zeppelin began touring in the United States, opening for Spirit," the suit states. "It was during this time that Jimmy Page, Led Zeppelin's guitarist, grew familiar with 'Taurus' and the rest of Spirit's catalog. Page stated in interviews that he found Spirit to be 'very good' and that the band's performances struck him 'on an emotional level.' "

- Suit claims similarities between two songs
- Randy California was guitarist for the group Spirit
- Jimmy Page has called the accusation "ridiculous"

**(Cloze-style) Query**

According to claims in the suit, "Parts of 'Stairway to Heaven,' instantly recognizable to the music fans across the world, sound almost identical to significant portions of '**X**.'"

**Reference Answers**

Taurus

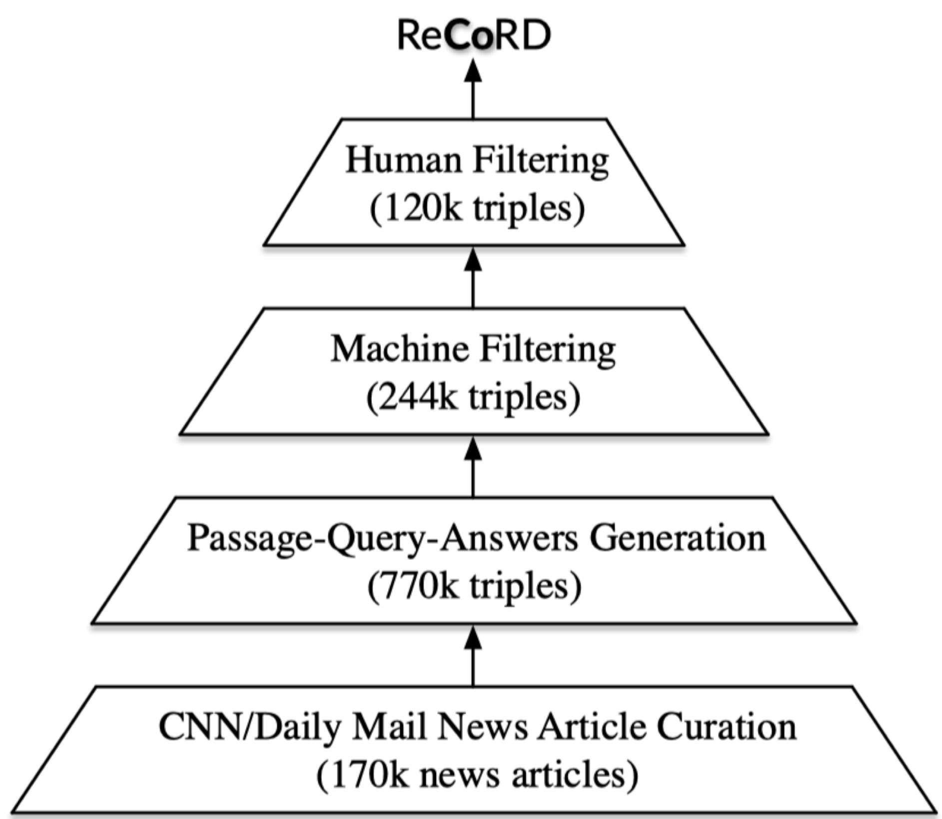



Figure 2: The overview of data collection stages.



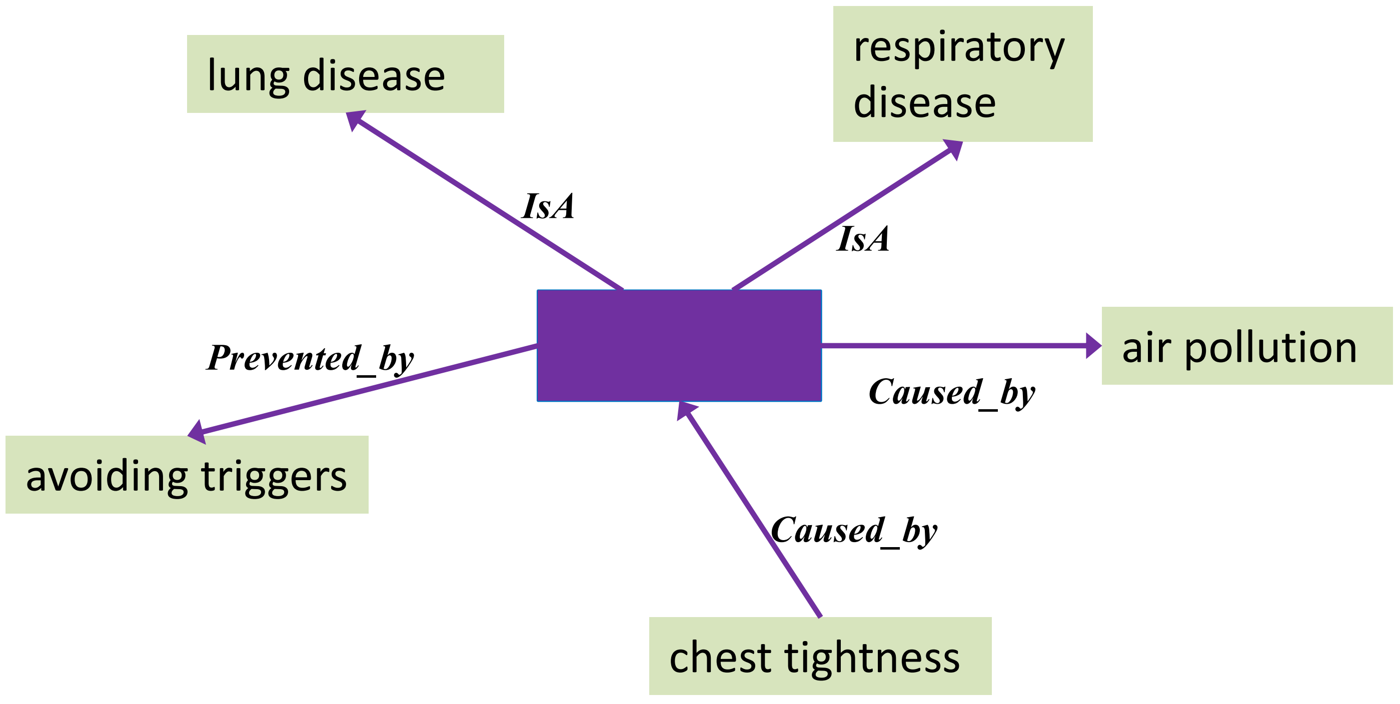
Zhang et al. 2018. ReCoRD: Bridging the Gap between Human Co and Machine Commonsense Reading Comprehension

# Knowledge
# in
# Dialog Generation

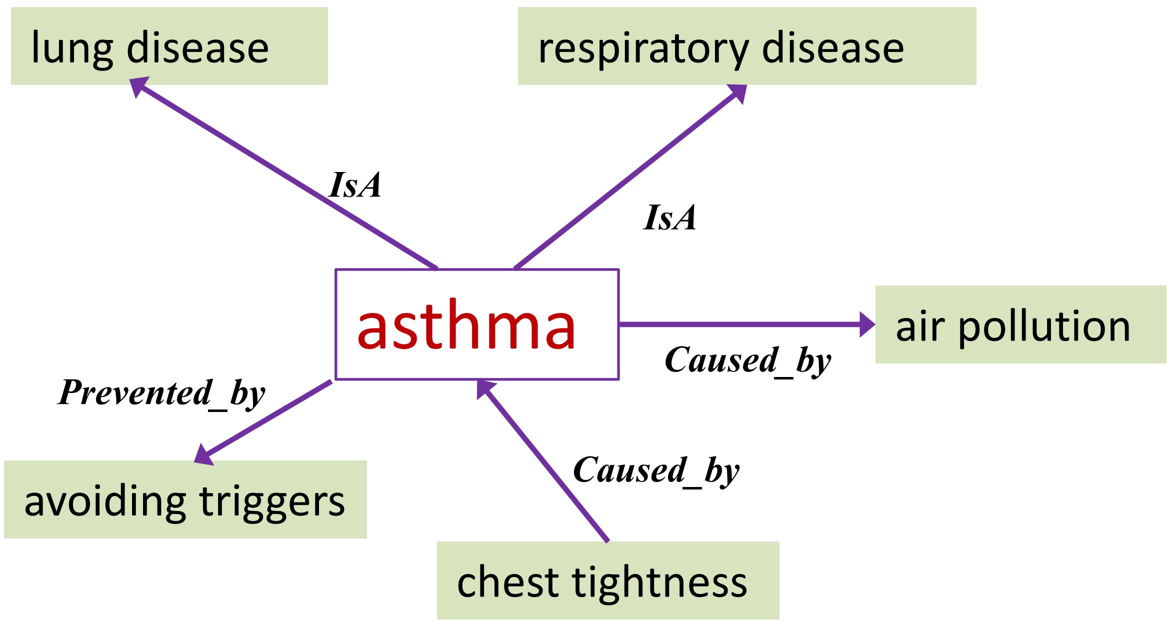
# Commonsense Knowledge

- **Commonsense knowledge** consists of facts about the everyday world, that all humans are expected to know. (Wikipedia)
  - Lemons are sour
  - Tree has leafs
  - Dog has four legs
- Commonsense Reasoning ~ **Winograd Schema Challenge**:
  - The trophy would not fit in the brown suitcase because it was too **big**. What was too **big**?
  - The trophy would not fit in the brown suitcase because it was too ***small***. What was too ***small***?

# Commonsense Knowledge



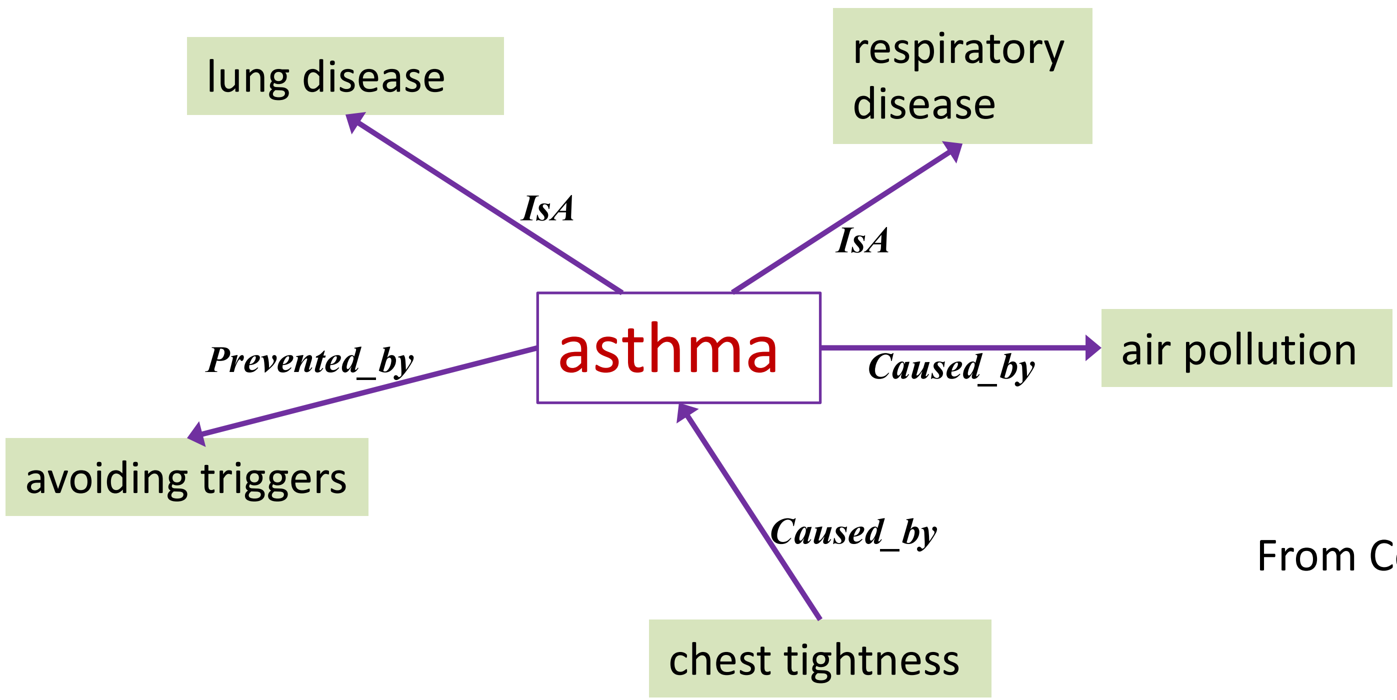
# Commonsense Knowledge



# Commonsense Knowledge

**Post: I have an asthma since three years old.**

**Triples in knowledge graph:**
**(lung disease, IsA, asthma )**
**(asthma, Prevented_by, avoiding triggers)**



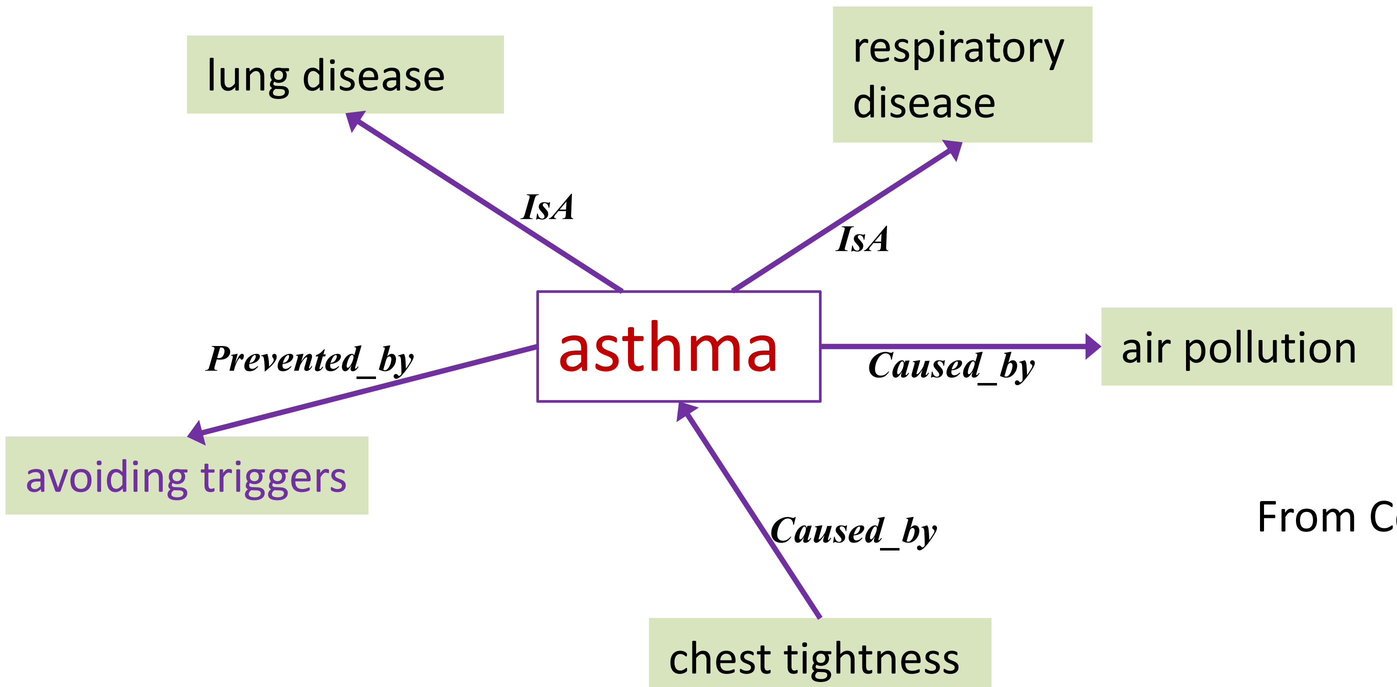

From ConceptNet

# Commonsense Knowledge in Chatbots

**Post: I have an asthma since three years old.**

**Triples in knowledge graph:**
**(lung disease, IsA, asthma )**
**(asthma, Prevented_by, avoiding triggers)**

**Response: I am sorry to hear that. Maybe avoiding triggers can prevent asthma attacks.**



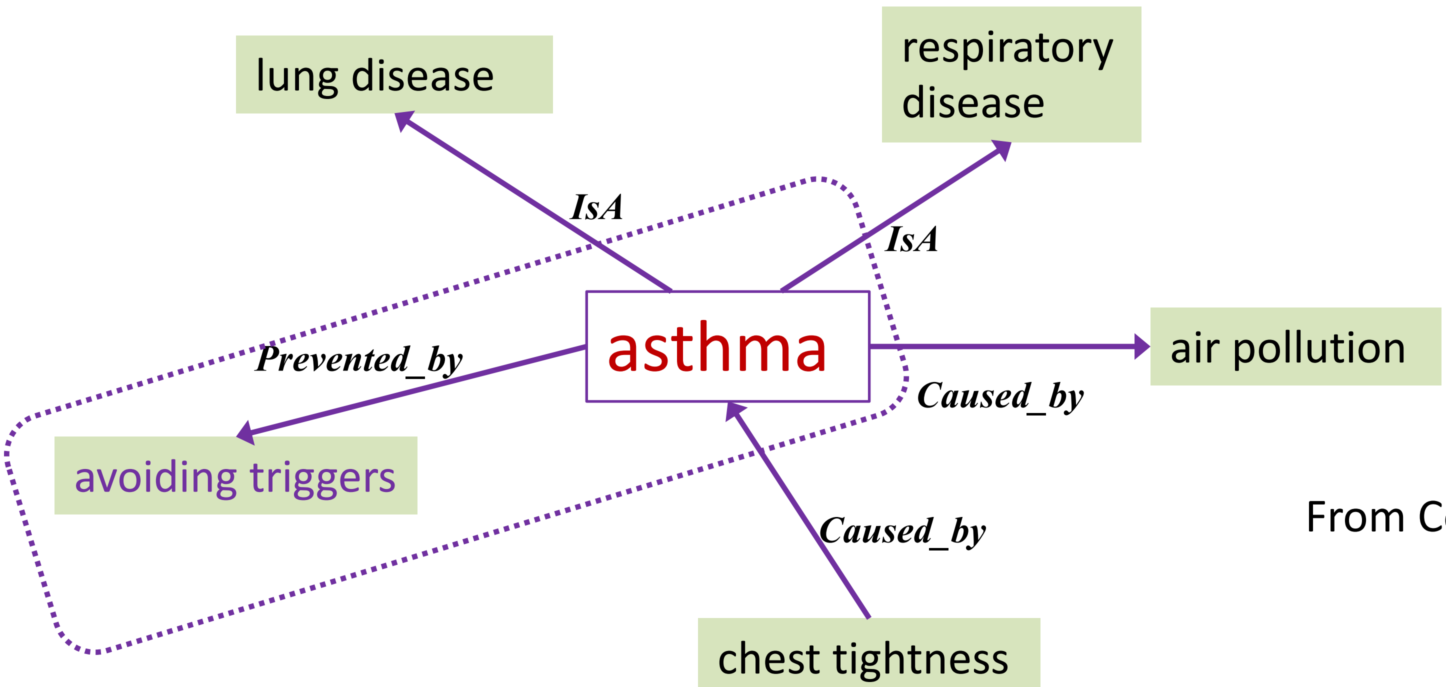

From ConceptNet

# Commonsense Knowledge in Chatbots

**Post: I have an asthma since three years old.**

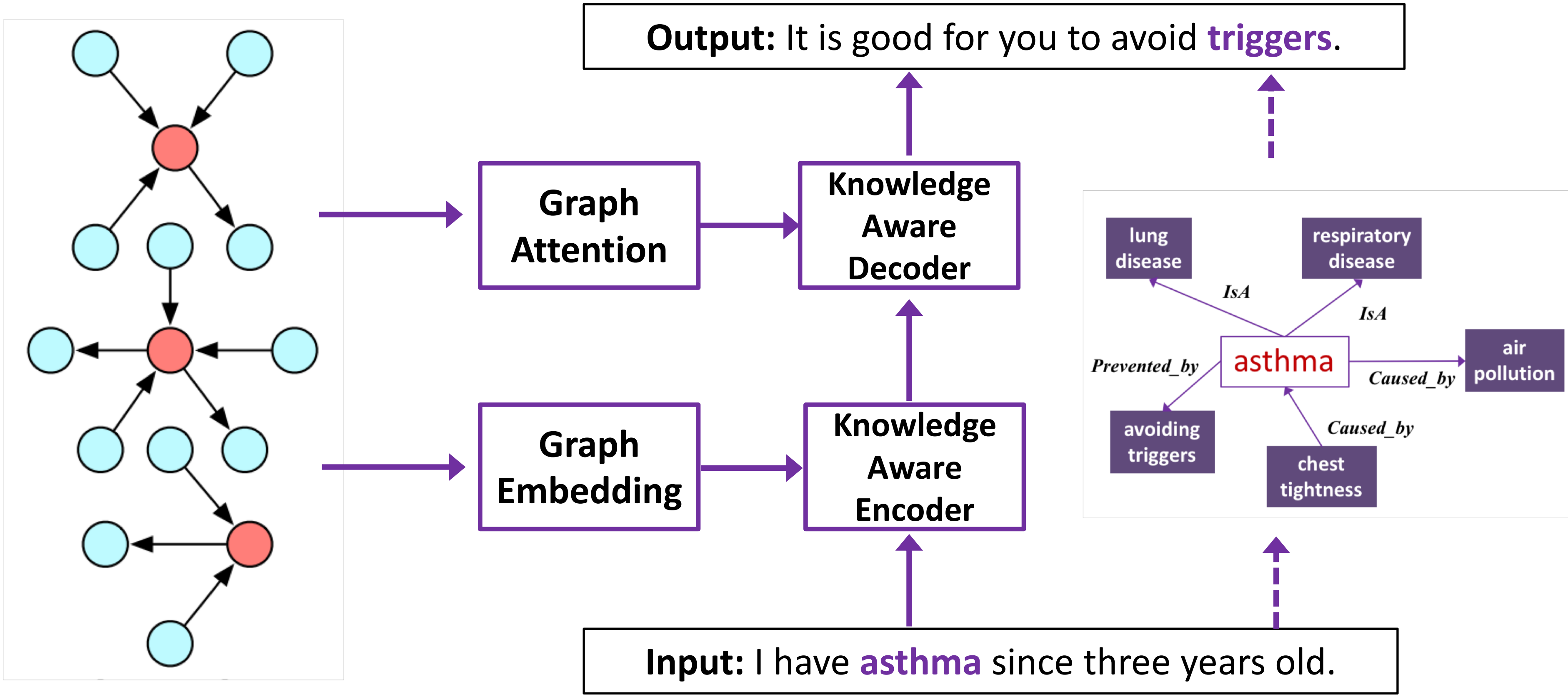
**Triples in knowledge graph:**
**(lung disease, IsA, asthma )**
**(asthma, Prevented_by, avoiding triggers)**

**Response: I am sorry to hear that. Maybe avoiding triggers can prevent asthma attacks.**



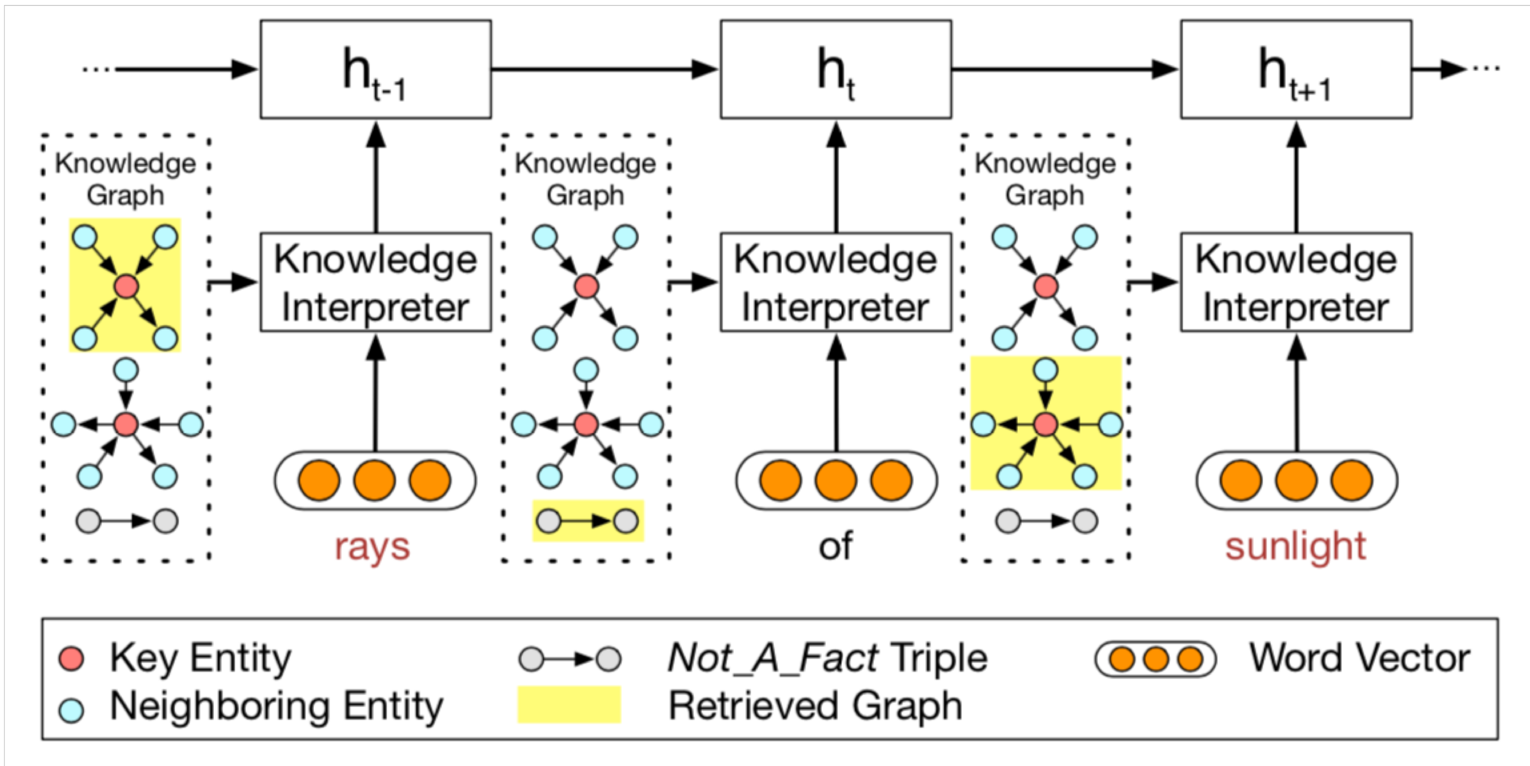

From ConceptNet

# Commonsense-aware Dialog Generation



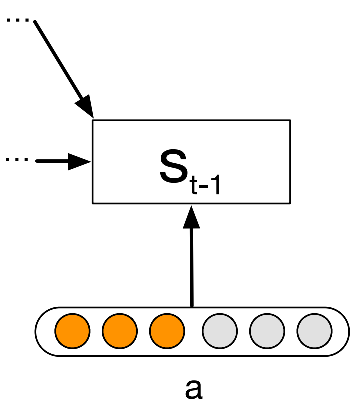
# Commonsense Knowledge in Chatbots

**Static graph attention**: encoding semantics in graph,
Feeding knowledge-enhanced info. into the encoder

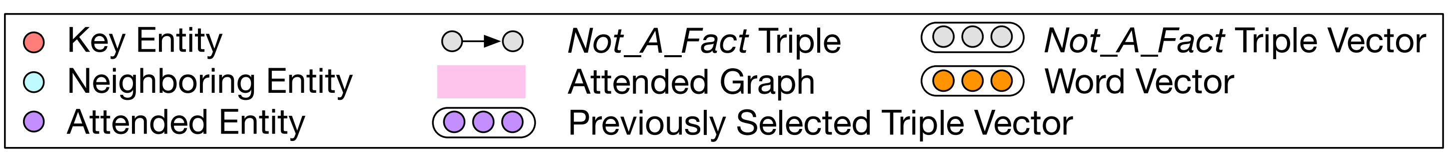


# Commonsense Knowledge in Chatbots

**Dynamic graph attention**: first attend a graph, then to a triple within that graph, finally generate with the words in a graph

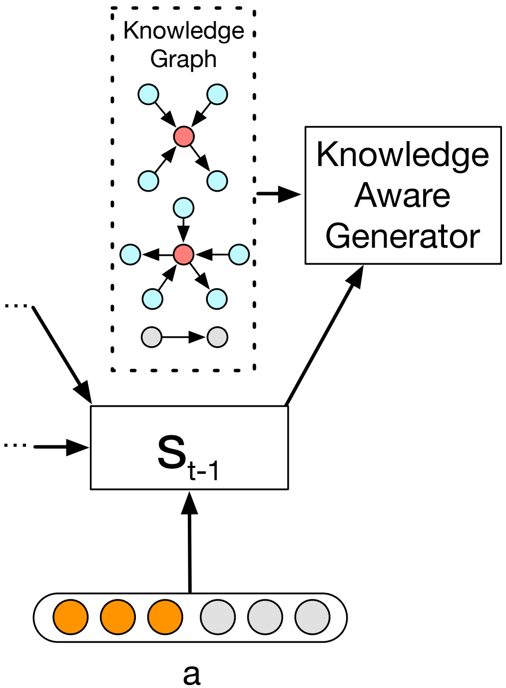



$$\boldsymbol{s}_{t+1} = \mathbf{GRU}(\boldsymbol{s}_t, [\boldsymbol{c}_t; \boldsymbol{c}_t^g; \boldsymbol{c}_t^k; \boldsymbol{e}(y_t)]),$$
$$\boldsymbol{e}(y_t) = [\boldsymbol{w}(y_t); \boldsymbol{k}_j],$$



# Commonsense Knowledge in Chatbots

**Dynamic graph attention**: first attend a graph, then to a triple within that graph, finally generate with the words in a graph

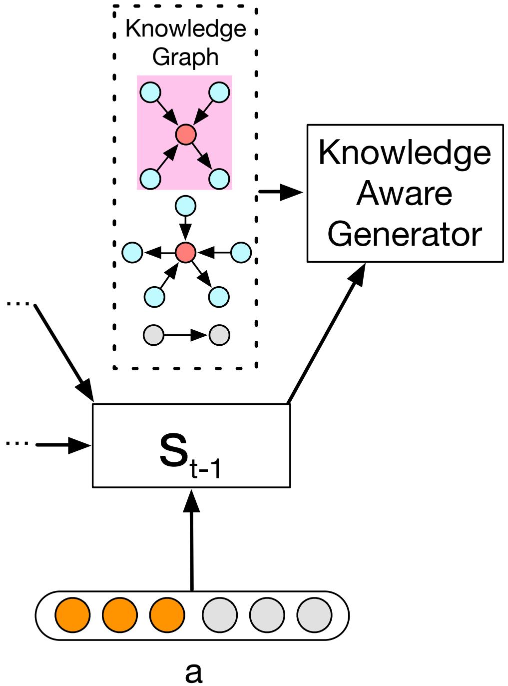



$$
\begin{aligned}
\boldsymbol{g}_i &= \sum_{n=1}^{N_{g_i}} \alpha_n^s [\boldsymbol{h}_n; \boldsymbol{t}_n], \\
\alpha_n^s &= \frac{\exp(\beta_n^s)}{\sum_{j=1}^{N_{g_i}} \exp(\beta_j^s)}, \\
\beta_n^s &= (\mathbf{W_r} \boldsymbol{r}_n)^\top \tanh(\mathbf{W_h} \boldsymbol{h}_n + \mathbf{W_t} \boldsymbol{t}_n),
\end{aligned}
$$

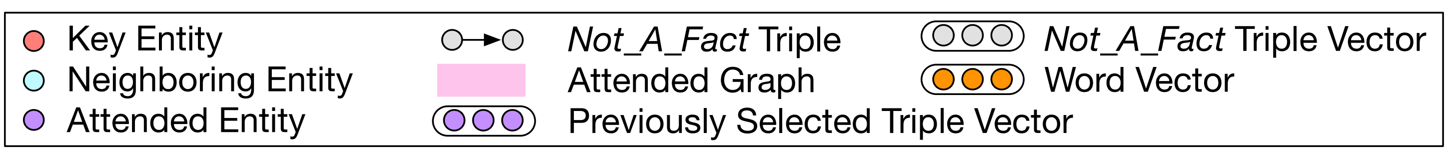
Key Entity
Neighboring Entity
Attended Entity
*Not_A_Fact* Triple
Attended Graph
Previously Selected Triple Vector
*Not_A_Fact* Triple Vector
Word Vector

# Commonsense Knowledge in Chatbots

**Dynamic graph attention**: first attend a graph, then to a triple within that graph, finally generate with the words in a graph

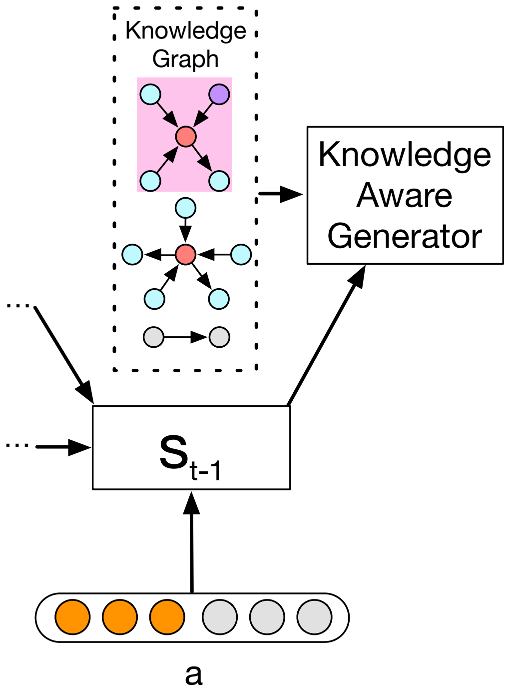



$$
\begin{aligned}
\boldsymbol{c}_t^g &= \sum_{i=1}^{N_G} \alpha_{ti}^g \boldsymbol{g}_i, \\
\alpha_{ti}^g &= \frac{\exp(\beta_{ti}^g)}{\sum_{j=1}^{N_G} \exp(\beta_{tj}^g)}, \\
\beta_{ti}^g &= \boldsymbol{V_b}^\top \tanh(\mathbf{W_b} \boldsymbol{s}_t + \mathbf{U_b} \boldsymbol{g}_i),
\end{aligned}
$$

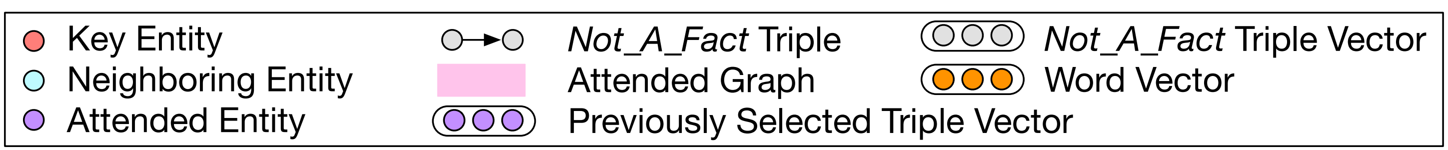


# Commonsense Knowledge in Chatbots

**Dynamic graph attention**: first attend a graph, then to a triple within that graph, finally generate with the words in a graph

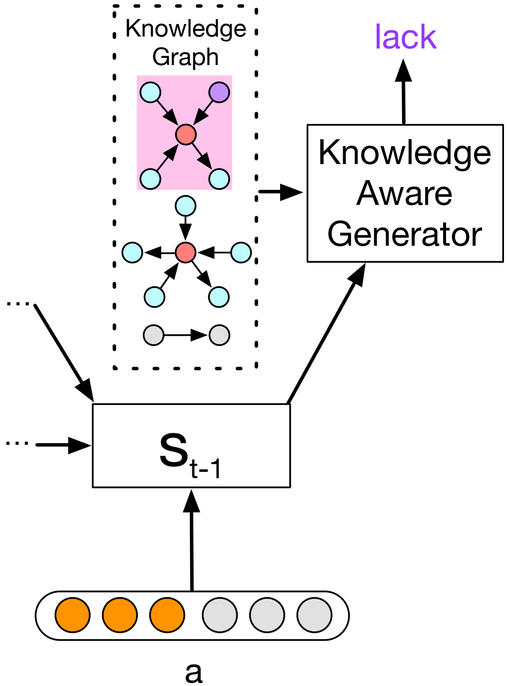



$$
\begin{aligned}
\boldsymbol{c}_t^k &= \sum_{i=1}^{N_G} \sum_{j=1}^{N_{g_i}} \alpha_{ti}^g \alpha_{tj}^k \boldsymbol{k}_j, \\
\alpha_{tj}^k &= \frac{\exp(\beta_{tj}^k)}{\sum_{n=1}^{N_{g_i}} \exp(\beta_{tn}^k)}, \\
\beta_{tj}^k &= \boldsymbol{k}_j^\top \mathbf{W_c} \boldsymbol{s}_t,
\end{aligned}
$$

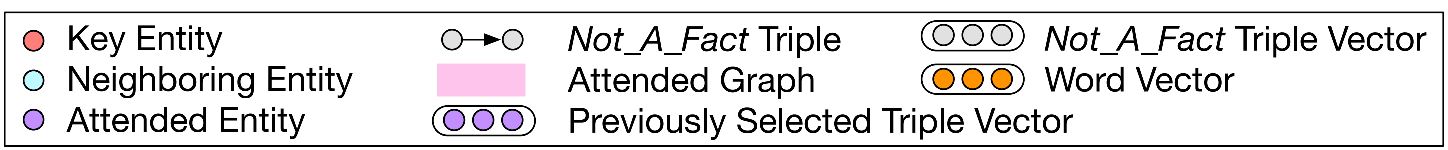


# Commonsense Knowledge in Chatbots

**Dynamic graph attention**: first attend a graph, then to a triple within that graph, finally generate with the words in a graph

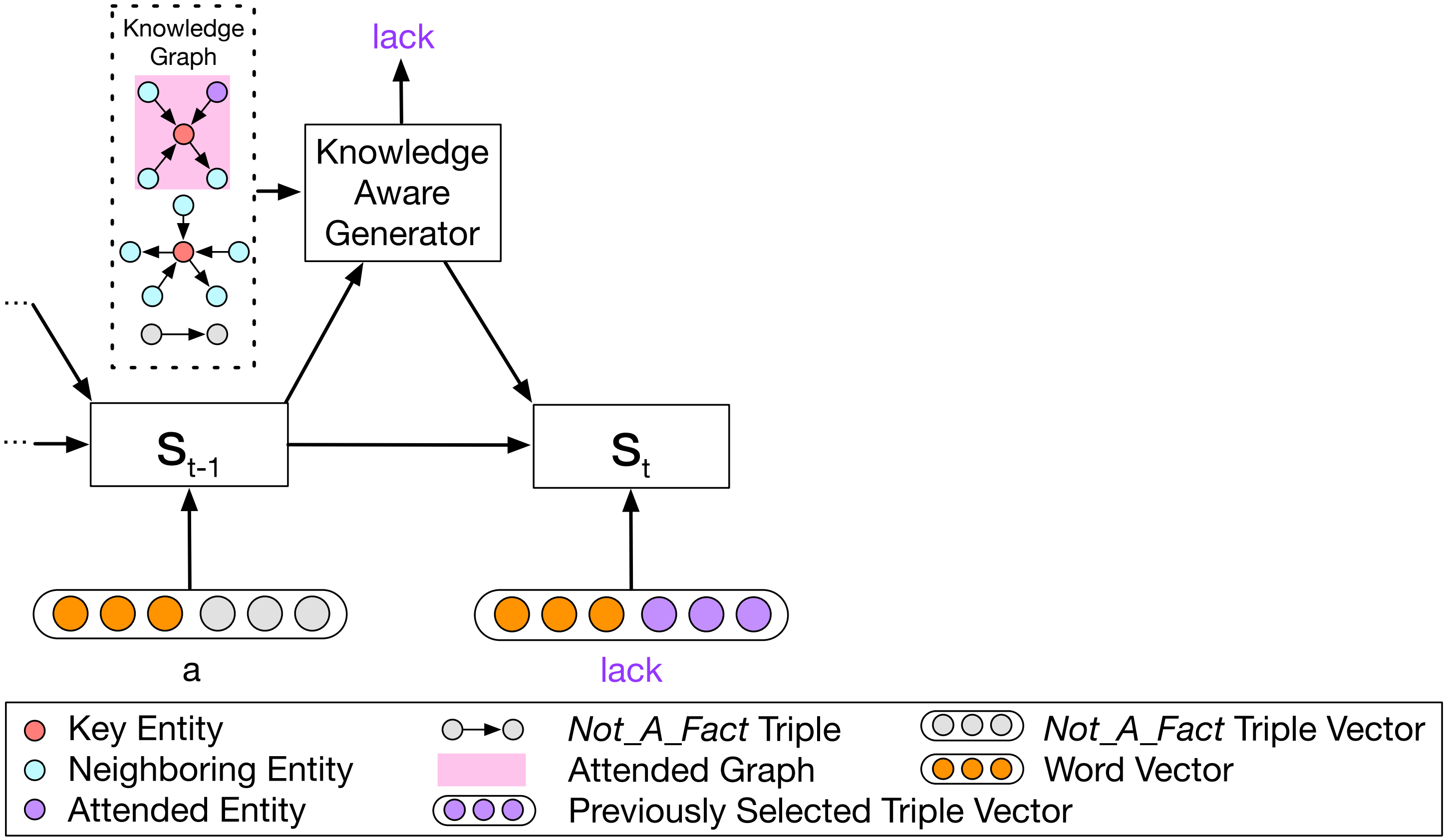



$$
\begin{aligned}
\boldsymbol{a}_t &= [\boldsymbol{s}_t; \boldsymbol{c}_t; \boldsymbol{c}_t^g; \boldsymbol{c}_t^k], \\
\gamma_t &= \mathrm{sigmoid}(\mathbf{V_o}^\top \boldsymbol{a}_t), \\
P_c(y_t = w_c) &= \mathrm{softmax}(\mathbf{W_o} \boldsymbol{a}_t), \\
P_e(y_t = w_e) &= \alpha_{ti}^g \alpha_{tj}^k, \\
y_t \sim \boldsymbol{o}_t = P(y_t) &= \begin{bmatrix} (1-\gamma_t) P_g(y_t = w_c) \\ \gamma_t P_e(y_t = w_e) \end{bmatrix},
\end{aligned}
$$



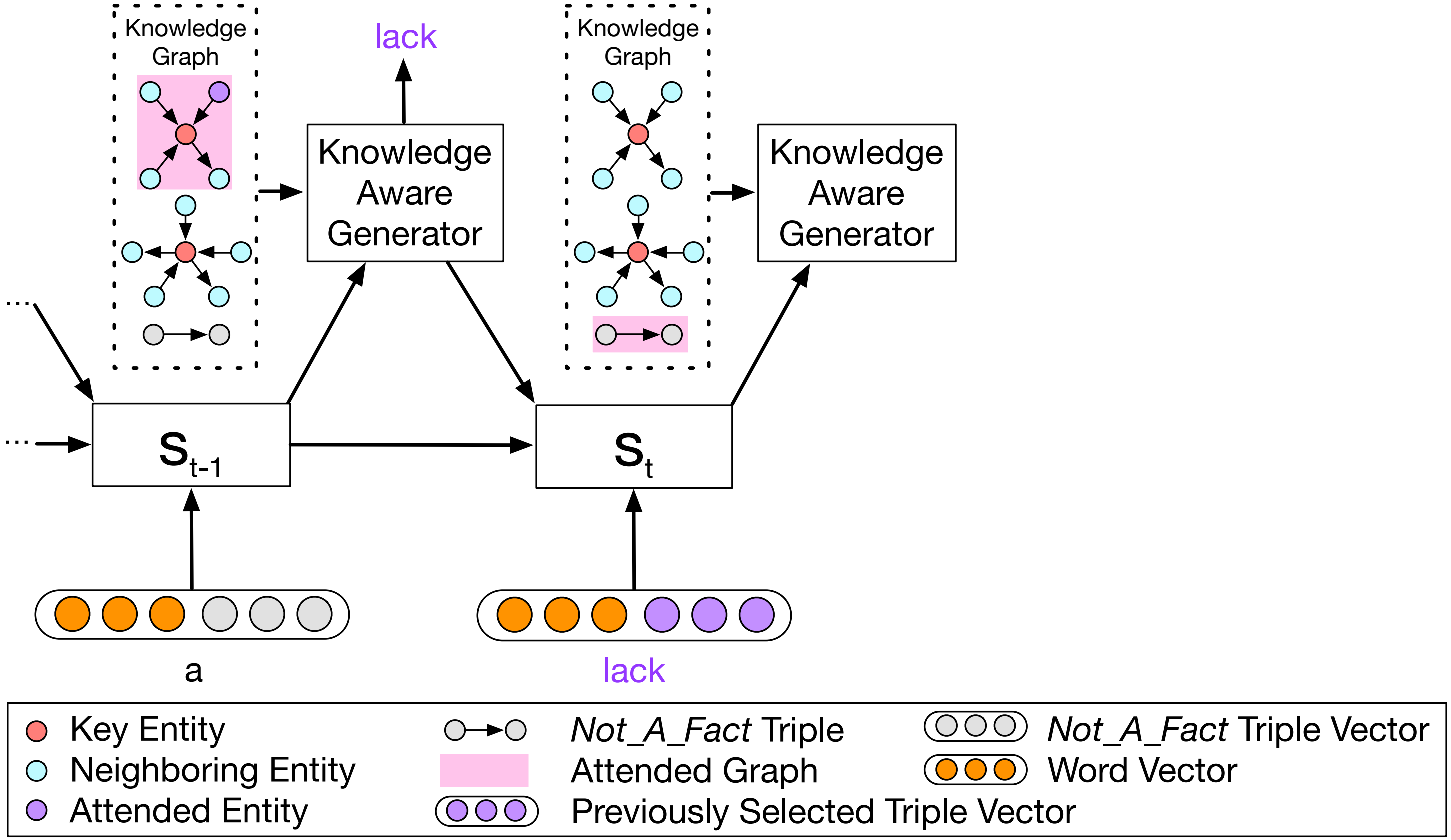
# Commonsense Knowledge in Chatbots

**Dynamic graph attention**: first attend a graph, then to a triple within that graph, finally generate with the words in a graph



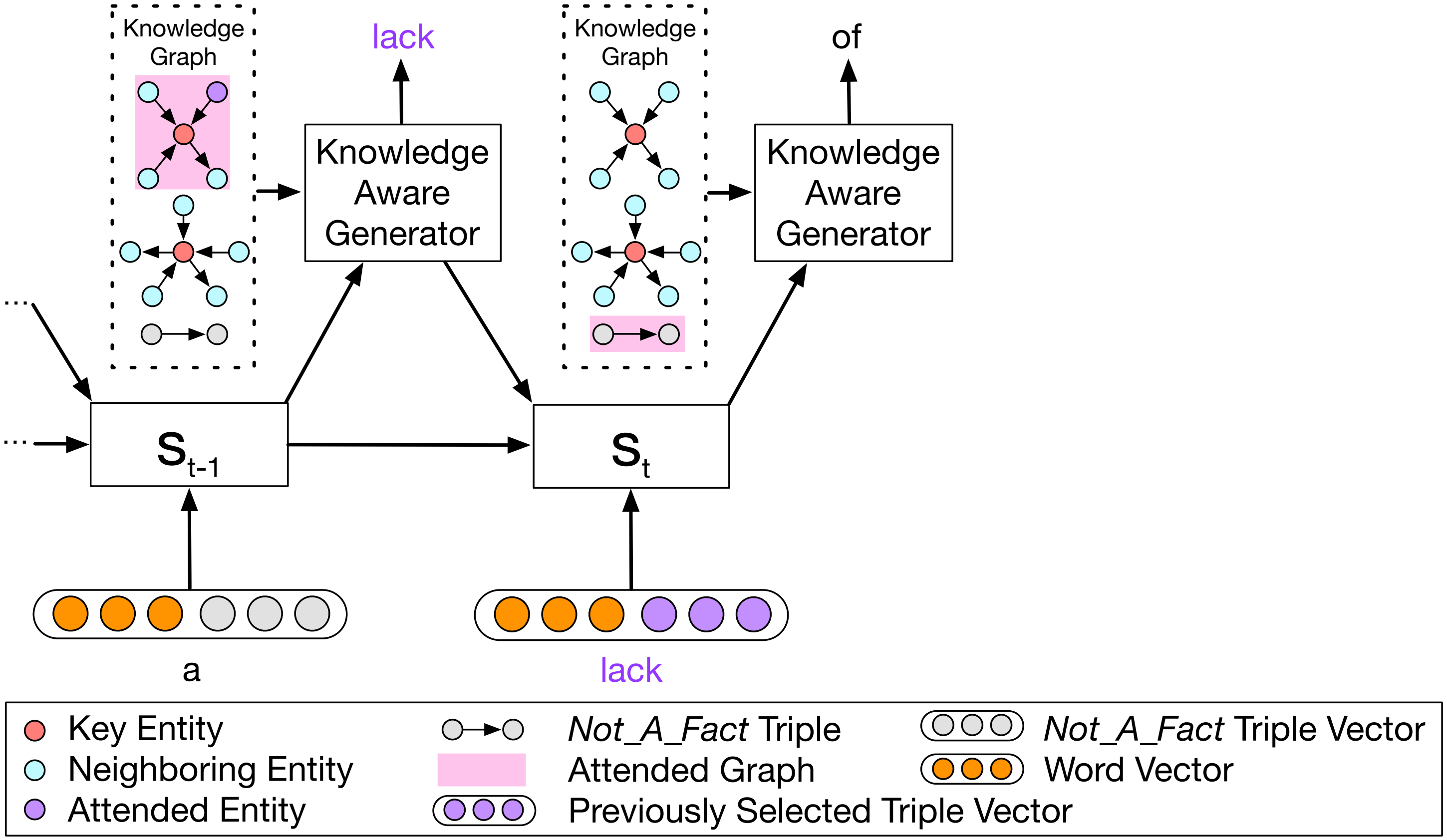
# Commonsense Knowledge in Chatbots

**Dynamic graph attention**: first attend a graph, then to a triple within that graph, finally generate with the words in a graph



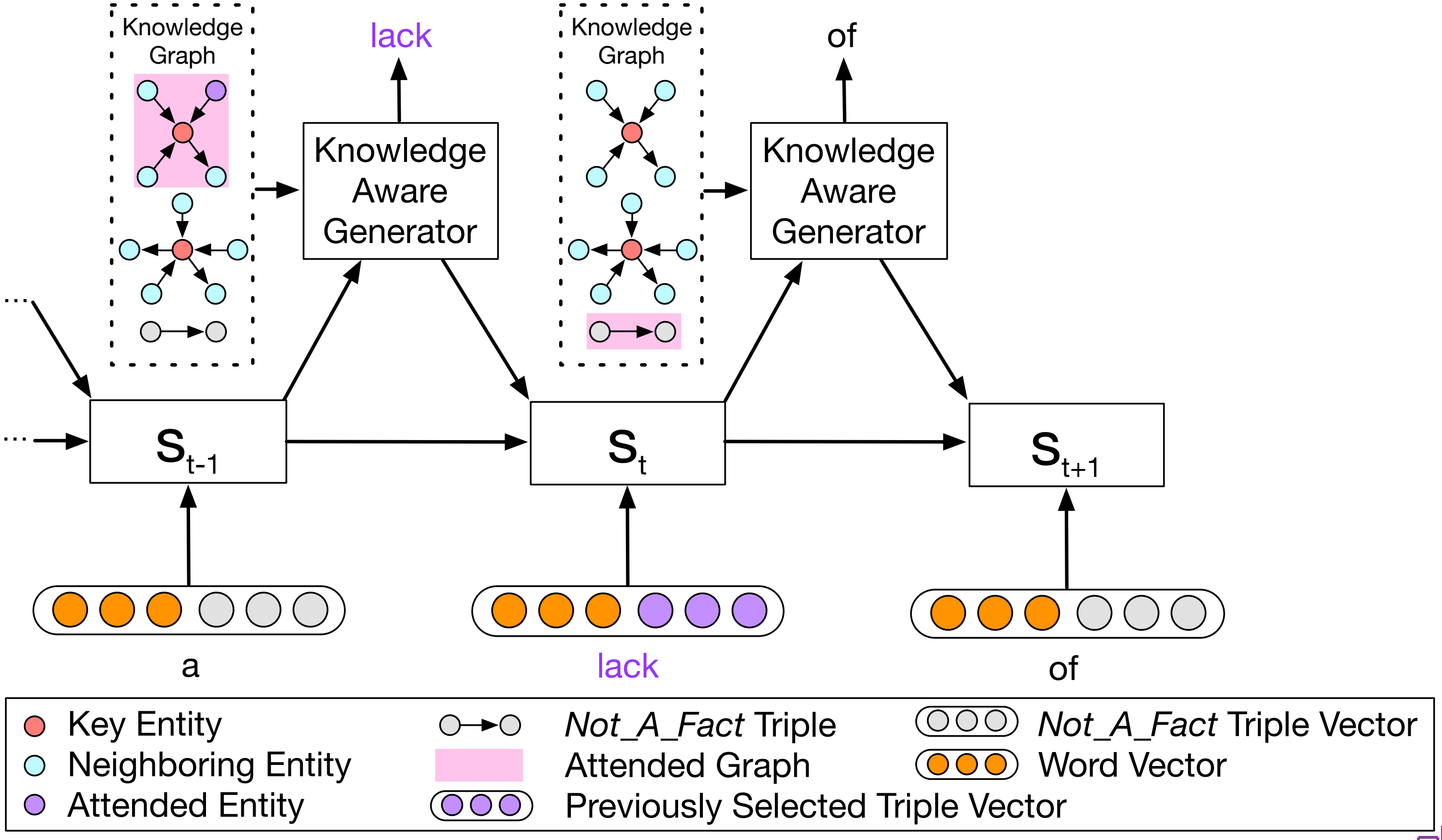
# Commonsense Knowledge in Chatbots

**Dynamic graph attention**: first attend a graph, then to a triple within that graph, finally generate with the words in a graph



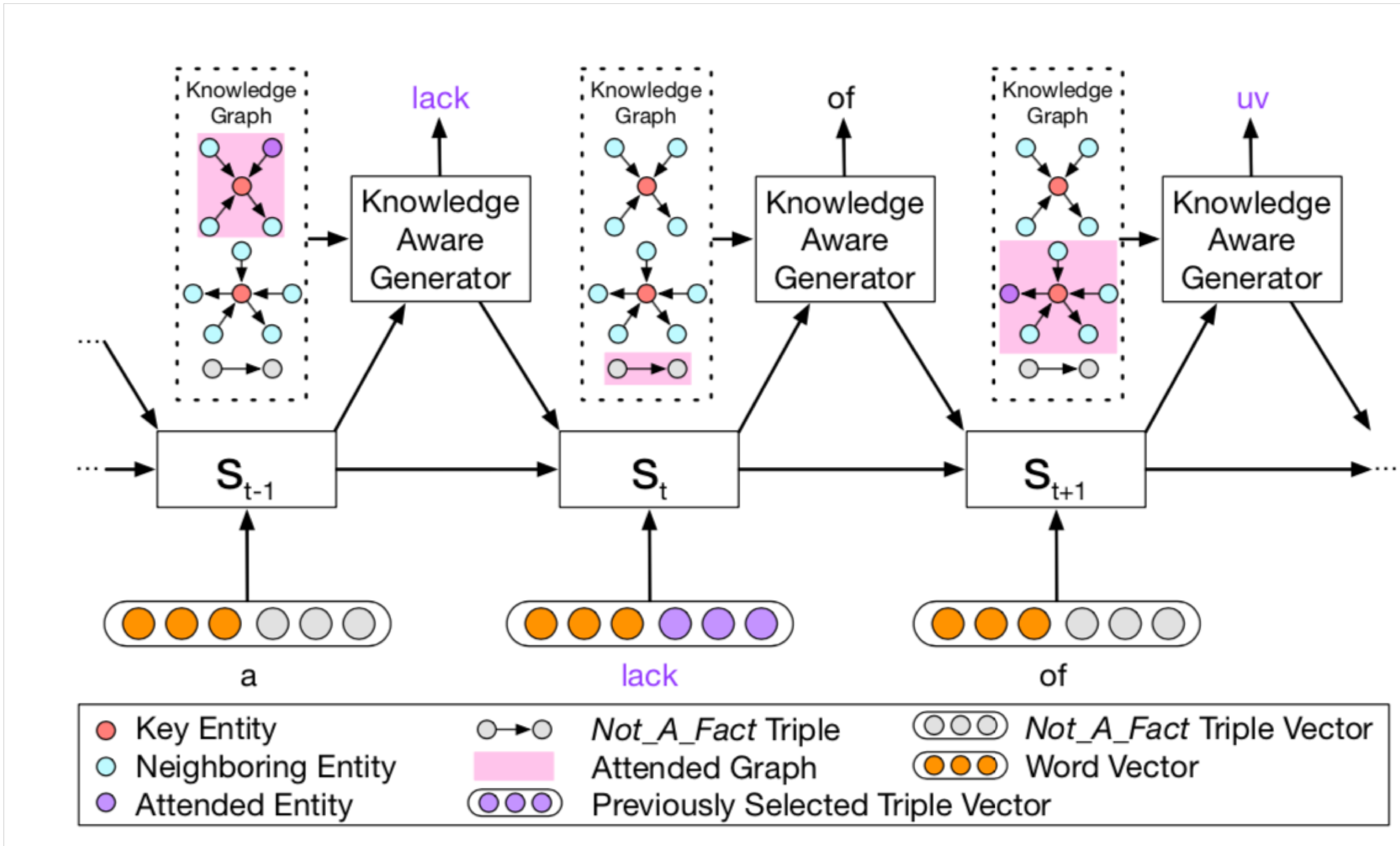
# Commonsense Knowledge in Chatbots

**Dynamic graph attention**: first attend a graph, then to a triple within that graph, finally generate with the words in a graph



# Commonsense Knowledge in Chatbots

**Dynamic graph attention**: first attend a graph, then to a triple within that graph, finally generate with the words in a graph



# Commonsense Knowledge in Chatbots

- Dataset: filtered from 10M reddit single-round dialogs

| Conversational Pairs | | Commonsense KB | |
|---|---|---|---|
| Training | 3,384,185 | Entity | 21,471 |
| Validation | 10,000 | Relation | 44 |
| Test | 20,000 | Triple | 120,850 |

Table 1: Statistics of the dataset and the knowledge base.

# Commonsense Knowledge in Chatbots

**Automatic evaluation**

| Model | Overall | | High Freq. | | Medium Freq. | | Low Freq. | | OOV | |
|---|---|---|---|---|---|---|---|---|---|---|
| | ppx. | ent. | ppx. | ent. | ppx. | ent. | ppx. | ent. | ppx. | ent. |
| Seq2Seq | 47.02 | 0.717 | 42.41 | 0.713 | 47.25 | 0.740 | 48.61 | 0.721 | 49.96 | 0.669 |
| MemNet | 46.85 | 0.761 | 41.93 | 0.764 | 47.32 | 0.788 | 48.86 | 0.760 | 49.52 | 0.706 |
| CopyNet | 40.27 | 0.96 | 36.26 | 0.91 | 40.99 | 0.97 | 42.09 | 0.96 | 42.24 | 0.96 |
| CCM | **39.18** | **1.180** | **35.36** | **1.156** | **39.64** | **1.191** | **40.67** | **1.196** | **40.87** | **1.162** |

**Manual evaluation** (Sign-test, p-value<0.005)

| Model | Overall | | High Freq. | | Medium Freq. | | Low Freq. | | OOV | |
|---|---|---|---|---|---|---|---|---|---|---|
| | app. | inf. | app. | inf. | app. | inf. | app. | inf. | app. | inf. |
| CCM vs. Seq2Seq | 0.616 | 0.662 | 0.605 | 0.656 | 0.549 | 0.624 | 0.636 | 0.650 | 0.673 | 0.716 |
| CCM vs. MemNet | 0.602 | 0.647 | 0.593 | 0.656 | 0.566 | 0.640 | 0.622 | 0.635 | 0.626 | 0.657 |
| CCM vs. CopyNet | 0.600 | 0.640 | 0.606 | 0.669 | 0.586 | 0.619 | 0.610 | 0.633 | 0.596 | 0.640 |

Hao Zhou, Tom Yang, Minlie Huang, Haizhou Zhao, Jingfang Xu, Xiaoyan Zhu. Commonsense Knowledge Aware Conversation Generation with Graph Attention. **IJCAI-ECAI 2018**, Stockholm, Sweden.

# Generation Examples 1

**Post**: He proposed March 5th. We will be married October 10th. So 7 months

**Knowledge**: (engaged, RelatedTo, married), (wedding, RelatedTo, married), (miss, DistinctFrom, married), (March, RelatedTo, months)

- **Seq2Seq**: I'm not sure how to answer that. I'm not sure how to make a date.
- **MemNet**: I'm assuming he's married to a gay marriage?
- **CopyNet**: I'm sorry, I'm not sure what you mean by that.
- **CCM**: I'm pretty sure he's going to be **engaged** in the next 6 **months**.

# Generation Examples 2

**Post**: Totally thought it was going to be doug's grave.
**Knowledge**: (thought, RelatedTo, thinking), (doug, HasProperty, dead), (dead, RelatedTo, grave), (grave, RelatedTo, death)

- **Seq2Seq**: I was thinking the same thing.
- **MemNet**: I thought it was a dead end.
- **CopyNet**: I think it's a bit of a stretch.
- **CCM**: I thought it was doug's **death**. I was **thinking** the same thing.

# Knowledge in
# Story Ending Generation

# Story Ending Generation

- Given a story context, conclude the story and complete the plot (**reasonable in logic and causality**)

**Context:** Today is Halloween .
Jack is so excited to go trick or treating tonight .
He is going to dress up like a monster .
The costume is real scary .

**Ending :** He hopes to get a lot of candy .

# Story Ending Generation

Generating a good ending requires:

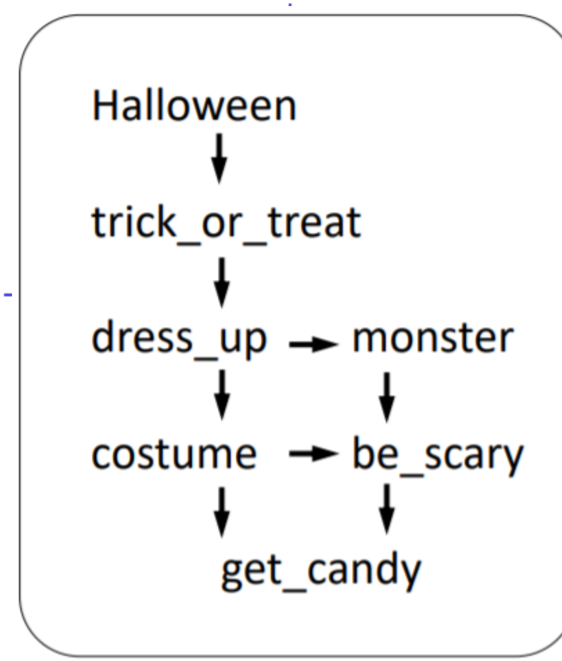
- Representing the **context clues** which contain key information for planning a reasonable ending
- Using **implicit knowledge** (e.g., commonsense knowledge) to facilitate understanding of the story and better predict what will happen next.

# Logic: Story Ending Generation

**Finding context clues:** plan the order of events and entities.

Today is **Halloween** .
Jack is so excited to go **trick or treating** tonight .
He is going to **dress up** like a **monster** .
The **costume** is real **scary** .

He hopes to get a lot of **candy** .



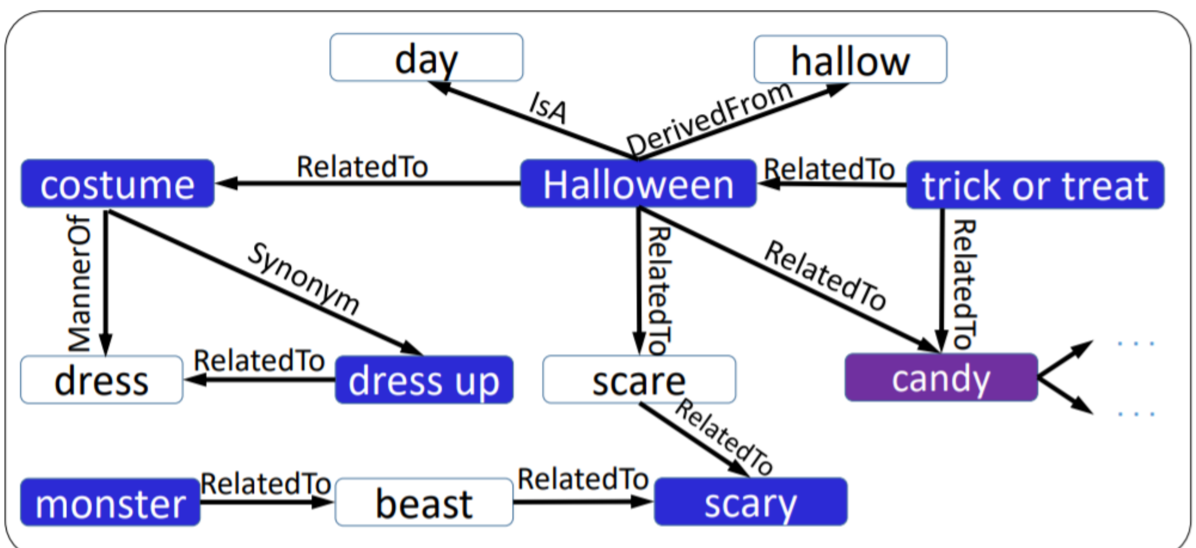

Jian Guan, Yansen Wang, Minlie Huang. **Story Ending Generation with Incremental Encoding and Commonsense Knowledge**. AAAI 2019

# Logic: Story Ending Generation

## Commonsense knowledge

Today is **Halloween** .
Jack is so excited to go **trick or treating** tonight .
He is going to **dress up** like a **monster** .
The **costume** is real **scary** .

He hopes to get a lot of **candy** .



# Task Overview

- Given a story context consisting of a sentence sequence:

$$X = \{X_1, X_2\,, X_2, \dots, X_K\}, \text{ where } X_i = x_1^{(i)} x_2^{(i)} \dots x_{l_i}^{(i)}$$

- The model should generate a one-sentence ending:

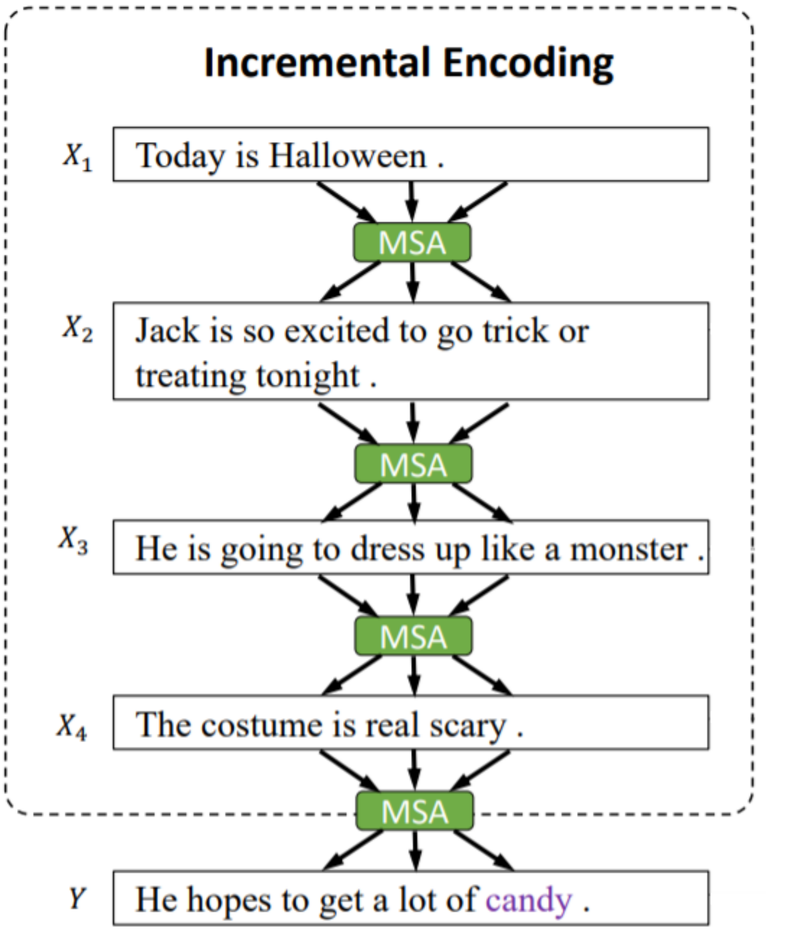
$$Y = y_1 y_2 \dots y_l$$

- Formally:

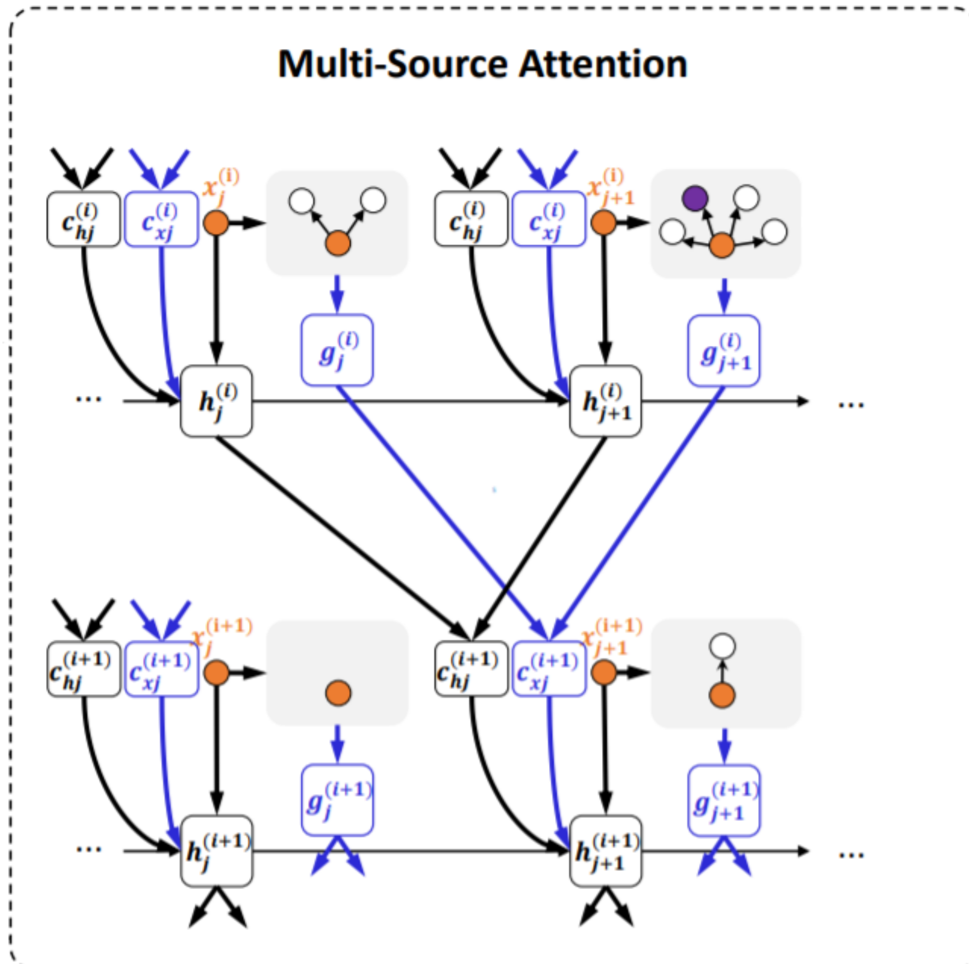
$$Y^* = \underset{Y}{argmax}\, \mathcal{P}(Y|X).$$

# Logic: Story Ending Generation

## Incremental Encoding



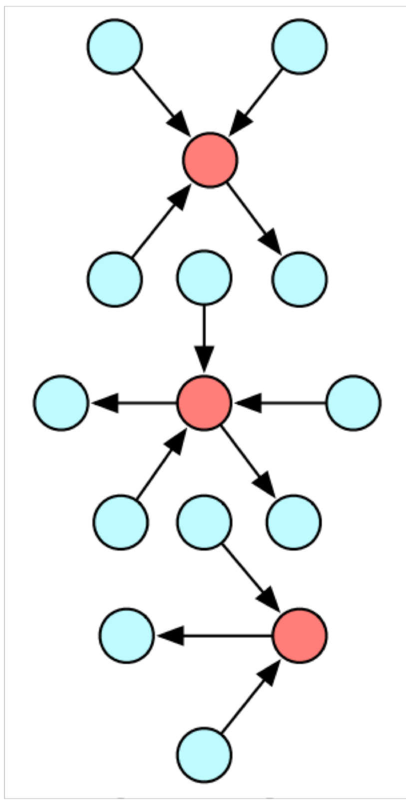

## Multi-Source Attention

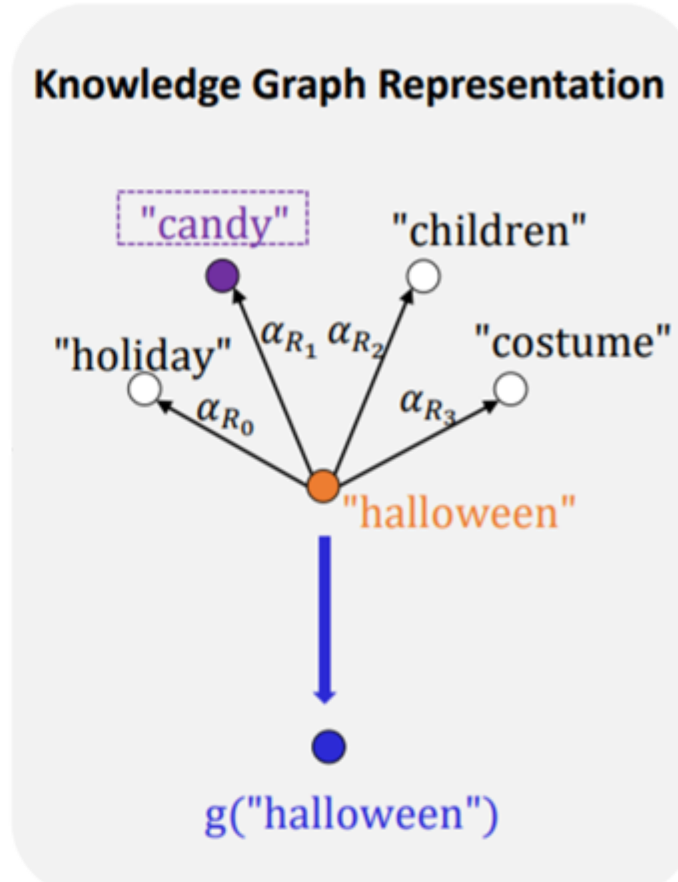


# Logic: Story Ending Generation

**Attention to the knowledge base:** static graph attention

**Graph attention**





# Model--- Encoder

Possible solutions for encoding:

- Concatenating the *K* sentences to a long sentence and encoding it with an LSTM
- Using a hierarchical LSTM with hierarchical attention (Yang et al. 2016)
- **Incremental Encoding (our proposal)**

# Model --- Encoder

Incremental Encoding

- Effective to represent the context clues which may **capture the key logic information.**
- The current sentence $X_i$
- An **attentive read of the preceding sentence** $X_{i-1}$: $\boldsymbol{c}_{\boldsymbol{l}j}^{(i)}$

$$\mathbf{h}_j^{(i)} = \mathbf{LSTM}(\mathbf{h}_{j-1}^{(i)}, e(x_j^{(i)}), \mathbf{c}_{\mathbf{l}j}^{(i)}),\ i \geq 2.$$

- Story ending generation:

$$\mathbf{s}_t = \mathbf{LSTM}(\mathbf{s}_{t-1}, \boldsymbol{e}(y_{t-1}), \mathbf{c}_{\mathbf{l}t}),$$

$$\mathcal{P}(y_t|y_{<t}, X) = \mathbf{softmax}(\mathbf{W}_0 \boldsymbol{s}_t + \mathbf{b}_0),$$

# Model ---Encoder

## Context vector

- Capture the relationship between words (or states) in the current sentence and those in the preceding sentence
- Encode implicit knowledge that is beyond the text
- Formally: $$\mathbf{c}_{\mathbf{l}j}^{(i)} = \mathbf{W}_{\mathbf{l}}([\mathbf{c}_{\mathbf{h}j}^{(i)}; \mathbf{c}_{\mathbf{x}j}^{(i)}]) + \mathbf{b}_{\mathbf{l}},$$
  - $\mathbf{c}_{\mathbf{h}j}^{(i)}$ is called **state context vector pointing to** $X_{i-1}$
  - $\mathbf{c}_{\mathbf{x}j}^{(i)}$ is called **knowledge context vector pointing to** $X_{i-1}$

# Model --- Encoder

- **State context vector**

$$\mathbf{c}_{\mathbf{h}j}^{(i)} = \sum_{k=1}^{l_{i-1}} \alpha_{h_k,j}^{(i)} \mathbf{h}_k^{(i-1)},$$

$$\alpha_{h_k,j}^{(i)} = \frac{e^{\beta_{h_k,j}^{(i)}}}{\sum_{m=1}^{l_{i-1}} e^{\beta_{h_m,j}^{(i)}}},$$

$$\beta_{h_k,j}^{(i)} = \mathbf{h}_{j-1}^{(i)\mathrm{T}} \mathbf{W_s} \mathbf{h}_k^{(i-1)},$$

- **Knowledge context vector**

$$\mathbf{c}_{\mathbf{x}j}^{(i)} = \sum_{k=1}^{l_{i-1}} \alpha_{x_k,j}^{(i)} \mathbf{g}(x_k^{(i-1)}),$$

$$\alpha_{x_k,j}^{(i)} = \frac{e^{\beta_{x_k,j}^{(i)}}}{\sum_{m=1}^{l_{i-1}} e^{\beta_{x_m,j}^{(i)}}},$$

$$\beta_{x_k,j}^{(i)} = \mathbf{h}_{j-1}^{(i)\mathrm{T}} \mathbf{W_k} \mathbf{g}(x_k^{(i-1)}),$$

# Model --- Knowledge

## Knowledge graph retrieval

- **ConceptNet**: a commonsense semantic network
- Consists of triples $\boldsymbol{R} = (\boldsymbol{h}, \boldsymbol{r}, \boldsymbol{t})$ meaning that head concept $\boldsymbol{h}$ has the relation $\boldsymbol{r}$ with tail concept $\boldsymbol{t}$
  - e.g. (*costume*, */R/MannerOf*, *dress*)
- Each word in a sentence is used as a query to **retrieve a one-hop graph** from ConceptNet.

# Model --- Knowledge

- The knowledge graph for a word extends (encodes) its meaning by **representing the graph** from neighboring concepts and relations.
  - **Graph Attention** (Velikovi et al. 2018; Zhou et al. 2018)
  - **Contextual attention** (Mihaylov and Frank 2018)

# Model --- Knowledge

- **Graph Attention**

$$\mathbf{g}(x) = \sum_{i=1}^{N_x} \alpha_{R_i} [\mathbf{h}_i ; \mathbf{t}_i],$$

$$\alpha_{R_i} = \frac{e^{\beta_{R_i}}}{\sum_{j=1}^{N_x} e^{\beta_{R_j}}},$$

$$\beta_{R_i} = (\mathbf{W_r r}_i)^{\mathrm{T}} \, tanh(\mathbf{W_h h}_i + \mathbf{W_t t}_i),$$

- **Contextual Attention**

$$\mathbf{g}(x) = \sum_{i=1}^{N_x} \alpha_{R_i} \mathbf{M}_{R_i},$$

$$\mathbf{M}_{R_i} = BiGRU(\mathbf{h}_i, \mathbf{r}_i, \mathbf{t}_i),$$

$$\alpha_{R_i} = \frac{e^{\beta_{R_i}}}{\sum_{j=1}^{N_x} e^{\beta_{R_j}}},$$

$$\beta_{R_i} = \mathbf{h}_{(x)}^{\mathrm{T}} \mathbf{W_c M}_{R_i},$$

# Model --- Knowledge

- Impose supervision on both the encoding network and decoding network

$$\Phi = \Phi_{en} + \Phi_{de}$$

$$\Phi_{en} = \sum_{i=2}^{K} \sum_{j=1}^{l_i} - \log \mathcal{P}(x_j^{(i)} = \tilde{x}_j^{(i)} | x_{<j}^{(i)}, X_{<i}),$$

$$\Phi_{de} = \sum_{t} - \log \mathcal{P}(y_t = \tilde{y}_t | y_{<t}, X),$$

# Datasets

- ROCStories corpus
  - Each story consists of **five sentences**, our task is to generate the ending given the first 4 sentence
  - 90,000 for training and 8,162 for evaluation
  - Average length of $X_1/X_2/X_3/X_4/Y$ is 8.9/9.9/10.1/10.0/10.5

- ConceptNet
  - Only retrieve the relations whose head entity and tail entity are **noun or verb**, meanwhile **both occurring in SCT**.
  - Retain at most 10 triples if there are too many for a word.
  - Average number of triples for each query word is 3.4

# Metrics

- **Automatic Evaluation**
  - Perplexity, BLEU-1 and BLEU-2
    - How well a model fits the data

- **Manual Evaluation**
  - Grammar (Gram.)
    - Score 2 : without any grammar errors
    - Score 1 : with a few errors but still understandable
    - Score 0 : with severe errors and incomprehensible
  - Logicality (Logic.)
    - Score 2 : totally reasonable endings
    - Score 1 : relevant but with some discrepancy
    - Score 2 : totally incompatible endings

# Results

| Model | PPL | BLEU-1 | BLEU-2 | Gram. | Logic. |
|---|---|---|---|---|---|
| Seq2Seq | 18.97 | 0.1864 | 0.0090 | 1.74 | 0.70 |
| HLSTM | 17.26 | 0.2459 | 0.0242 | 1.57 | 0.84 |
| HLSTM+Copy | 19.93 | 0.2469 | 0.0248 | 1.66 | 0.90 |
| HLSTM+MSA(GA) | 15.75 | 0.2588 | 0.0253 | 1.70 | 1.06 |
| HLSTM+MSA(CA) | 12.53 | 0.2514 | 0.0271 | 1.72 | 1.02 |
| IE (ours) | 11.04 | 0.2514 | 0.0263 | **1.84** | 1.10 |
| IE+MSA(GA) (ours) | 9.72 | 0.2566 | 0.0284 | 1.68 | **1.26** |
| IE+MSA(CA) (ours) | **8.79** | **0.2682** | **0.0327** | 1.66 | 1.24 |

Table 1: Automatic and manual evaluation results.

# Examples

| | |
|---|---|
| **Context:** | Martha is **cooking** a special **meal** for her family. She **wants everything to be just right** for when they eat. Martha **perfects everything** and puts her **dinner** into the **oven**. Martha goes to **lay down** for a quick **nap**. |
| **Golden Ending:** | She **oversleeps** and runs into the **kitchen** to take out her **burnt dinner**. |
| **Seq2Seq:** | She was so happy to have a ***new cake***. |
| **HLSTM:** | Her family ***and her family*** are very happy with her **food**. |
| **HLSTM+ Copy:** | **Martha** is happy to be able to ***eat her family***. |
| **HLSTM+ GA:** | She is happy to be able to **cook her dinner**. |
| **HLSTM+ CA:** | She is very happy that she has made a new **cook** . |
| **IE:** | She is very happy with her **family**. |
| **IE+GA:** | When she gets back to the **kitchen**, she sees a **burning light** on the **stove**. |
| **IE+CA:** | She realizes the **food** and is happy she was ready to **cook** . |

# Generation Examples

**Story 1:**

**Context:**

Taj has never drank an espresso drink.
He ordered one while out with his friends.
The shot of espresso tasted terrible to him.
Taj found that he couldn't stop talking or moving.

**Generated Ending:**

He decided to never drink again.

**Story 2:**

**Context:**

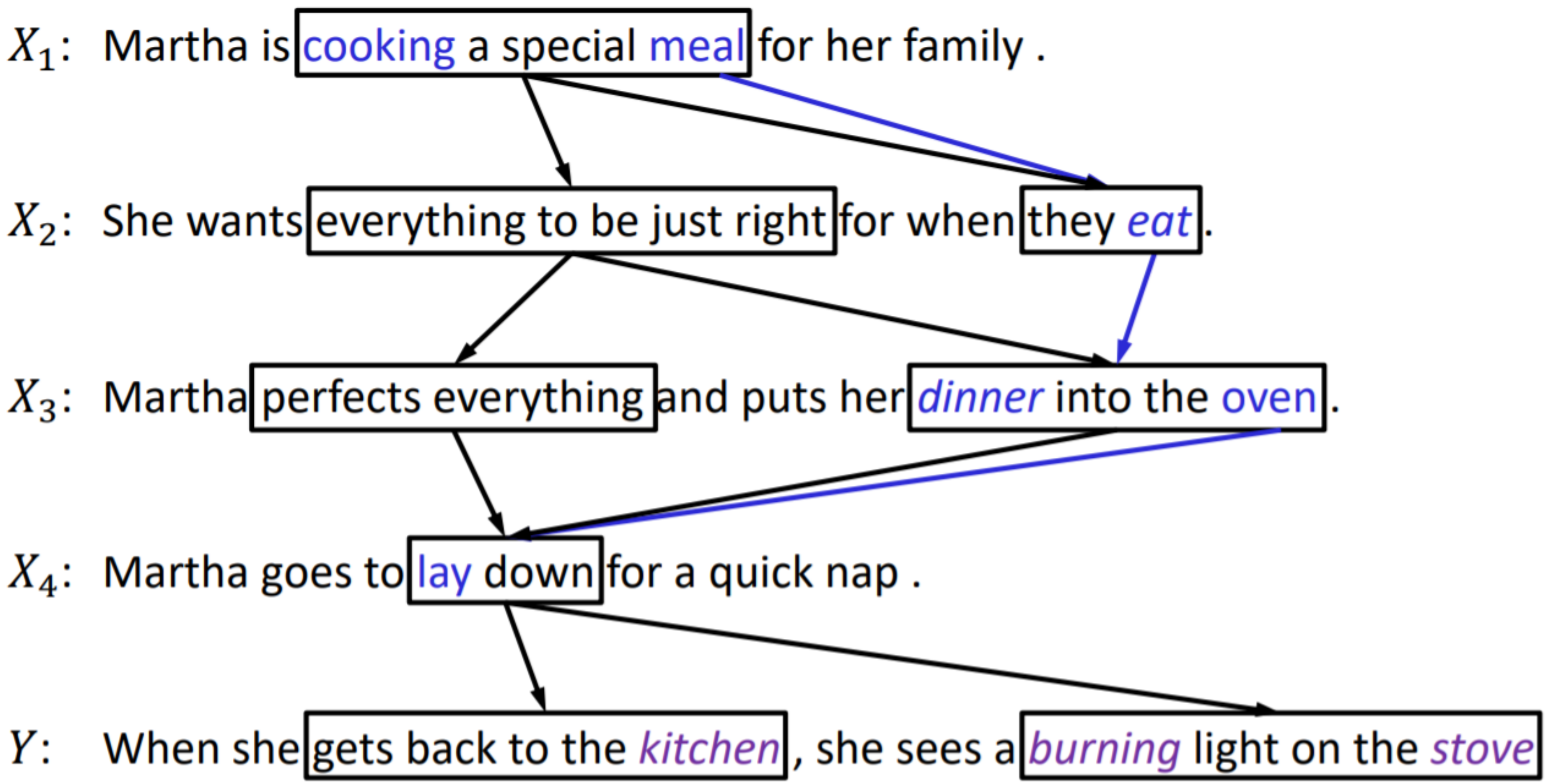
Martha is cooking a special meal for her family.
She wants everything to be just right for when they eat.
Martha perfects everything and puts her dinner into the oven.
Martha goes to lay down for a quick nap.

**Generated Ending:**

When she gets back to the kitchen, she sees a burning light on the stove.

# "Logic Chains": Contextual Clue

## Building context clues incrementally



# Knowledge-grounded Dialog Generation

| | |
|---|---|
| **Name** | The Shape of Water |
| **Year** | 2017 |
| **Director** | Guillermo del Toro |
| **Genre** | Fantasy, Drama |
| **Cast** | Sally Hawkins as Elisa Esposito, a mute cleaner who works at a secret government laboratory.<br>Michael Shannon as Colonel Richard Strickland, a corrupt military official,<br>Richard Jenkins as Giles, Elisa's closeted neighbor and close friend who is a struggling advertising illustrator.<br>Octavia Spencer as Zelda Delilah Fuller, Elisa's co-worker and friend who serves as her interpreter.,<br>Michael Stuhlbarg as Dimitri Mosenkov, a Soviet spy working as a scientist studying the creature, under the alias Dr. Robert Hoffstetler. |

A Dataset for Document Grounded Conversations (Zhou et al., 2018)

# Knowledge-grounded Dialog Generation

| | |
|---|---|
| User 2: | I thought The Shape of Water was one of Del Toro's best works. What about you? |
| User 1: | Did you like the movie? |
| User 1: | Yes, his style really extended the story. |
| User 2: | I agree. He has a way with fantasy elements that really helped this story be truly beautiful. |

Table 2: An example conversation for scenario 2. Both User 1 and User 2 have access to the Wiki document. The full dialogue is attached in the Appendix.

A Dataset for Document Grounded Conversations (Zhou et al., 2018)

# Knowledge-grounded Dialog Generation

| | |
|---|---|
| **Topic:** | Lifeguard |
| Apprentice: | So I am a lifeguard. Know anything about saving lives in water? |
| Wizard: | I'm impressed! It's a big responsibility to supervise other people's safety in the water! Tell me more. |
| Apprentice: | Well, I help make sure people do not drown or get injured while in or near the water! |
| **Knowledge:** | A lifeguard is a rescuer who supervises the safety and rescue of swimmers, surfers, . . .<br>Lifeguards are strong swimmers and trained in CPR/AED first aid, certified in water . . .<br>. . .<br>In some areas, the lifeguard service also carries out mountain rescues, or may function as the primary EMS provider. |
| Wizard: | I've heard that in some places, lifeguards also help with other sorts of emergencies, like mountain rescues!<br>Is that part of your job too? |
| Apprentice: | I have! I feel like you know much about this! What brings you to know so much? |
| Wizard: | Oh, that's about the extent of my knowledge. I've just been around beaches and I've always admired lifeguards. I'm not a super strong swimmer myself. |

Wizard of Wikipedia: Knowledge-Powered Conversational agents. Dinan et al. 2018.

# Summary

# Thanks for Your Attention

- http://coai.cs.tsinghua.edu.cn/ds/ 对话系统技术平台
- Acknowledgements
  - Prof Xiaoyan Zhu, Tsinghua colleagues, collaborators
  - Our students
- Contact:
  - Minlie Huang, Tsinghua University
  - aihuang@tsinghua.edu.cn
  - http://coai.cs.tsinghua.edu.cn/hml