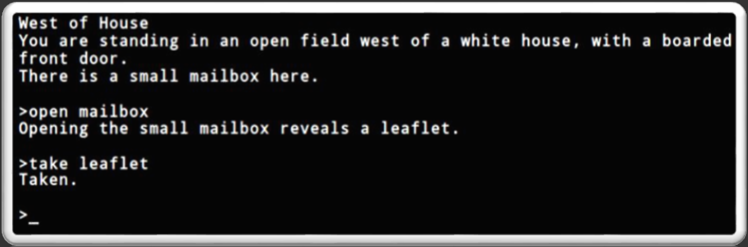
>>> WordPlay:
>>> Natural Language Understanding Through Text-based Game Solving

Name: Ronen Tamari (Hyadata Data Science Lab, HUJI)
Date: April 9, 2019



>>> Objective

Understand this...

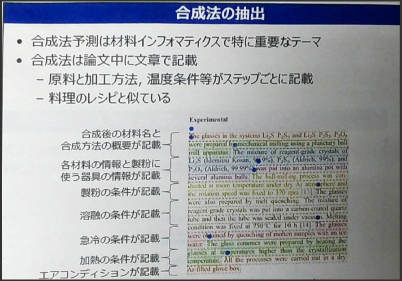
"The glasses in the systems $Li_2S-P_2S_5$ and $Li_2S-P_2S_5-P_2O_5$ were prepared by mechanical milling using a planetary ball mill apparatus. The mixture of reagent-grade crystals of Li2S (Idemitsu Kosan, 99.9%), P_2S_5 (Aldrich, 99%), and P_2O_5 (Aldrich, 99.99%) was put into an alumina pot with several alumina balls. The ball-milling process was conducted at room temperature under dry Ar atmosphere and the rotation speed was fixed to 370 rpm. The glasses were also prepared by melt quenching. The mixture of reagent-grade crystals was put into a carbon-coated quartz tube and then the tube was sealed under vacuum. Melting condition was fixed at 750°C for 10 h. The glasses were obtained by quenching of molten samples with an ice water. The glass ceramics were prepared by heating the glasses at temperatures higher than the crystallization temperature. All the processes were carried out in a dry Ar-filled glove box."

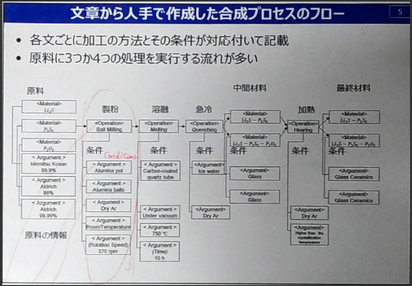
>>> Objective

* Action-graph extraction:

>>> Objective

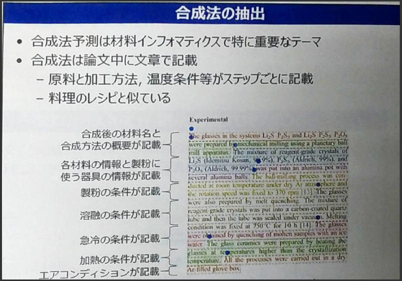
* Action-graph extraction:

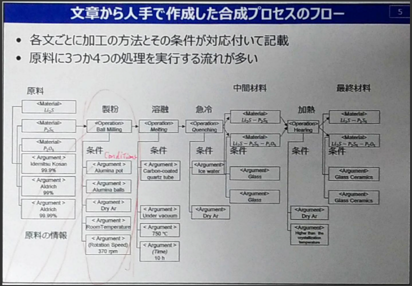




>>> Objective

* Action-graph extraction:

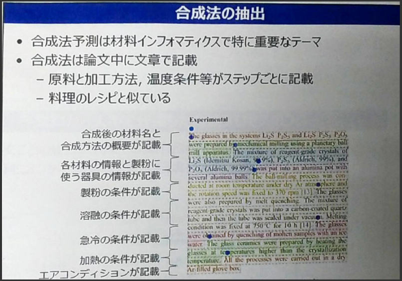


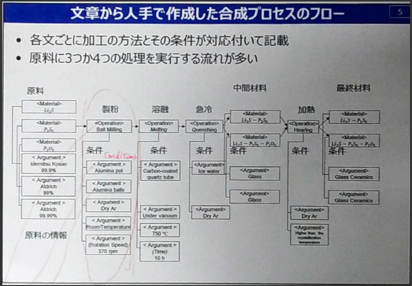


* Mat. sci. literature contains *millions* of such synthesis routes described in unstructured natural language texts.

>>> Objective

* Action-graph extraction:





* Mat. sci. literature contains *millions* of such synthesis routes described in unstructured natural language texts.
* Building a structured knowledge base from them remains a grand challenge in materials science.

>>> Problem: NLP is brittle[1]

And we don't even get the simple stuff...

[1] Based on https://newgeneralization.github.io/slides/YejinChoi.pdf

>>> Problem: NLP is brittle[1]

And we don't even get the simple stuff...

* Language models (LMs) are passive learners → One doesn't learn to write just by reading.

[1] Based on https://newgeneralization.github.io/slides/YejinChoi.pdf

>>> Problem: NLP is brittle[1]

And we don't even get the simple stuff...

* Language models (LMs) are passive learners $\rightarrow$ One doesn't learn to write just by reading.
* Language models are surface learners $\rightarrow$ Lack underlying world model.

[1] Based on https://newgeneralization.github.io/slides/YejinChoi.pdf

>>> Problem: NLP is brittle[1]

And we don't even get the simple stuff...

* Language models (LMs) are passive learners → One doesn't learn to write just by reading.
* Language models are surface learners → Lack underlying world model.
  "He was dressed in a white t-shirt, blue jeans, and a black t-shirt."

[1] Based on https://newgeneralization.github.io/slides/YejinChoi.pdf

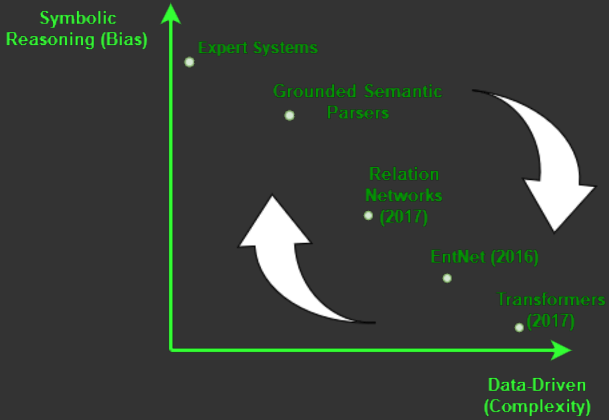
>>> Problem: NLP is brittle[1]

And we don't even get the simple stuff...

* Language models (LMs) are passive learners → One doesn't learn to write just by reading.
* Language models are surface learners → Lack underlying world model.
  "He was dressed in a white t-shirt, blue jeans, and a black t-shirt."
* Specialized procedure understanding models currently require volumes of fine-grained annotated data for good performance.

[1] Based on https://newgeneralization.github.io/slides/YejinChoi.pdf

>>> Surveying The Landscape



>>> Surveying The Landscape

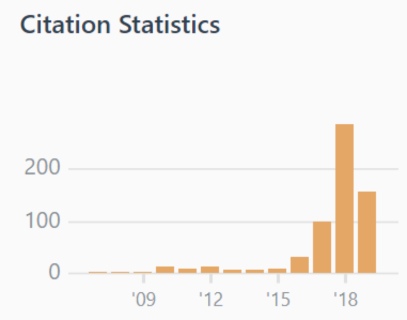


Figure: Graph Neural Network paper[4] citations per year, since publication in 2009 (from https://www.semanticscholar.org/, April 2019)

>>> Challenges

>>> Challenges

* Annotated data extremely limited.

>>> Challenges

* Annotated data extremely limited.
* Manual annotation is fine-grained, costly and reliant on domain-specific knowledge.

>>> Challenges

* Annotated data extremely limited.
* Manual annotation is fine-grained, costly and reliant on domain-specific knowledge.
* Highly technical language $\rightarrow$ transfer-learning limited in usefulness.

\>>> Spoiler...

>>> Spoiler...

* We didn't solve the problem yet...

>>> Spoiler...

* We didn't solve the problem yet...
* What we are doing is working a novel approach for this and (hopefully) many other similar problems, with some potentially interesting new extensions.

>>> Overview

>>> Understanding by Simulation - (White) Elephant In The Room?

>>> Understanding by Simulation - (White) Elephant In The Room?

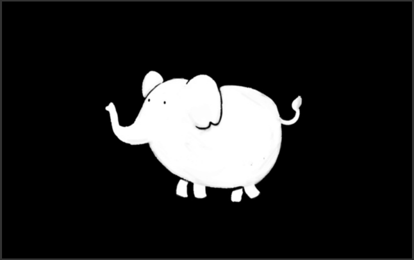
* Meaning is far more than just knowing syntax and word definitions!

>>> Understanding by Simulation - (White) Elephant In The Room?

* Meaning is far more than just knowing syntax and word definitions!
* When we process language, our mind constructs rich mental simulations, and this is an integral part of understanding.

>>> Understanding by Simulation - (White) Elephant In The Room?

* Meaning is far more than just knowing syntax and word definitions!
* When we process language, our mind constructs rich mental simulations, and this is an integral part of understanding.



\>>> Image Schema & Mental Spaces - Building Blocks of Human Cognition?

>>> Image Schema & Mental Spaces - Building Blocks of Human Cognition?

Image Schema[2]: *"Dynamic analog structures arising from perception, bodily movements, manipulation of objects, and experience of force..."*

* CONTAINMENT: "A text document contains words."
* SUPPORT: "We meet on Thursday."
* PATH, LINK, PART-OF etc...
* Limited vocabulary, but can be compositionally combined to express complex concepts!

>>> Image Schema & Mental Spaces - Building Blocks of Human Cognition?

>>> Image Schema & Mental Spaces - Building Blocks of Human Cognition?

Mental Spaces: *"A medium for conceptualization and thought... Any fixed or ongoing state of affairs as we conceptualize it is represented by a mental space."*

\>>> Image Schema & Mental Spaces - Building Blocks of Human Cognition?

Mental Spaces: *"A medium for conceptualization and thought... Any fixed or ongoing state of affairs as we conceptualize it is represented by a mental space."*

* Contains "mental entities".

\>>> Image Schema & Mental Spaces - Building Blocks of Human Cognition?

Mental Spaces: *"A medium for conceptualization and thought... Any fixed or ongoing state of affairs as we conceptualize it is represented by a mental space."*

* Contains "mental entities".
* Spaces are extendable, in that additional entities may be added to them in the course of cognitive processing.

>>> Image Schema & Mental Spaces - Building Blocks of Human Cognition?

Mental Spaces: *"A medium for conceptualization and thought... Any fixed or ongoing state of affairs as we conceptualize it is represented by a mental space."*

* Contains "mental entities".
* Spaces are extendable, in that additional entities may be added to them in the course of cognitive processing.
* Structured by cognitive models (~rules)

>>> And What of Machine Cognition?

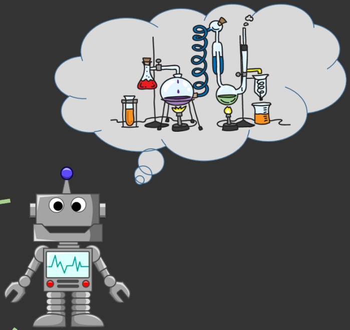
>>> And What of Machine Cognition?

* To what extent are such constructs necessary to understand language?

>>> And What of Machine Cognition?

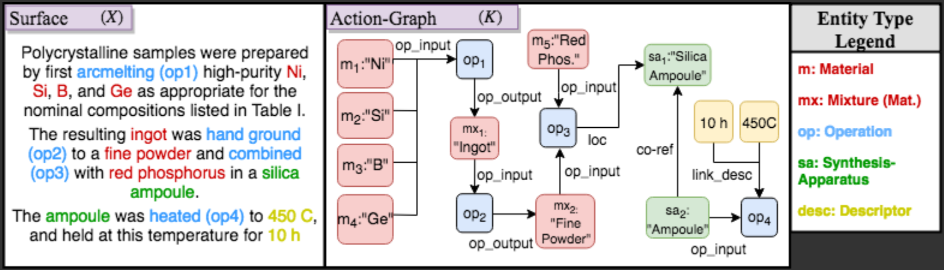
* To what extent are such constructs necessary to understand language?
* How could we endow a machine with such an inductive bias and how would it affect its abilities[6]?

"Polycrystalline samples were prepared by first arcmelting high-purity Ni, Si, B, and Ge as appropriate for the nominal compositions listed in Table I. The resulting ingot was hand ground to a fine powder and combined with red phosphorus in a silica ampoule that was subsequently sealed under vacuum. The ampoule was heated to 450 C, and held at this temperature for 10 h."



>>> Problem Framing

The underlying objective is *narrative sketch / action-graph extraction.*



>>> What's in a story?

Surface (X):

>>> What's in a story?

Surface (X):

* Natural language text assumed to have underlying narrative representation.
* Roughly a paragraph in our domain.

>>> What's in a story?

"Open" Action Graphs:

>>> What's in a story?

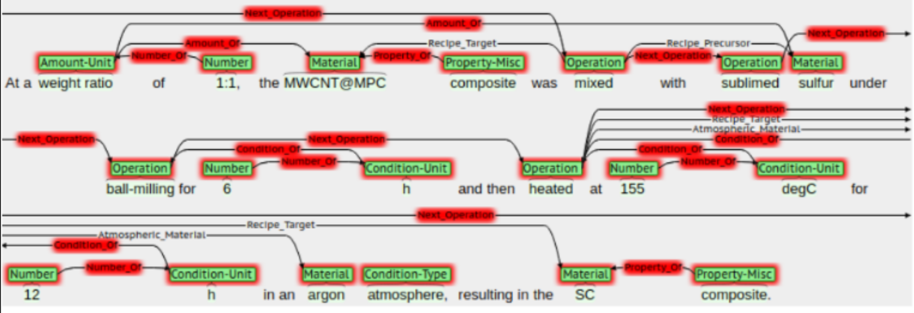
"Open" Action Graphs:

* Assume a vocabulary of entity and relation types $\mathcal{E} = \{e_1, ..., e_N\}$, $\mathcal{R} = \{r_1, ..., r_K\}$.

>>> What's in a story?

"Open" Action Graphs:

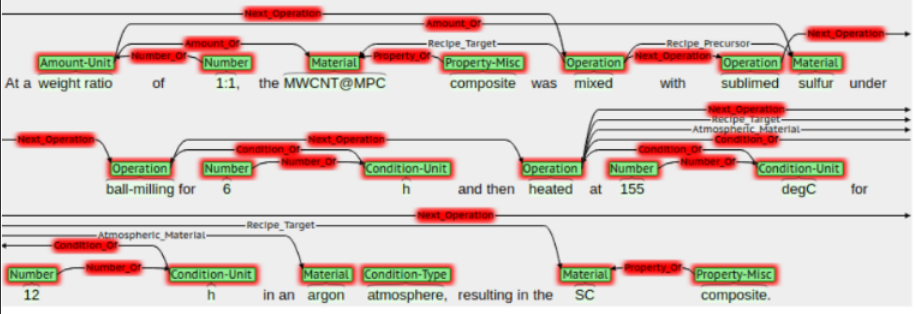
* Assume a vocabulary of entity and relation types $\mathcal{E} = \{e_1, ..., e_N\}$, $\mathcal{R} = \{r_1, ..., r_K\}$.



>>> What's in a story?

"Open" Action Graphs:

* Assume a vocabulary of entity and relation types $\mathcal{E} = \{e_1, ..., e_N\}$, $\mathcal{R} = \{r_1, ..., r_K\}$.



* Where can we add "mental simulation"?

Grounded Action Graph (K):

>>> What's in a story?

Grounded Action Graph (K):

* A *state* s is a set of facts over entities $\mathcal{E}$ and relations $\mathcal{R}$.

>>> What's in a story?

Grounded Action Graph (K):

* A *state* s is a set of facts over entities $\mathcal{E}$ and relations $\mathcal{R}$.
* World model Λ defined using linear logic. Defines legal relations between entities. A transition rule λ produces a new state from a given state:

>>> What's in a story?

Grounded Action Graph (K):

* A *state* s is a set of facts over entities $\mathcal{E}$ and relations $\mathcal{R}$.
* World model Λ defined using linear logic. Defines legal relations between entities. A transition rule λ produces a new state from a given state:

```
melt/p :: $at(P, r) & $at(fd, r) & $in(m, I) & powder(m) -> liquid(m)
```

>>> What's in a story?

Grounded Action Graph (K):

* A *state* s is a set of facts over entities $\mathcal{E}$ and relations $\mathcal{R}$.
* World model Λ defined using linear logic. Defines legal relations between entities. A transition rule λ produces a new state from a given state:

```
 melt/p :: $at(P, r) & $at(fd, r) & $in(m, I) & powder(m) ->
liquid(m)
```

Facts	~	Image Schema
States	~	Mental Spaces
World Model	~	Cognitive Model

>>> What's in a story?

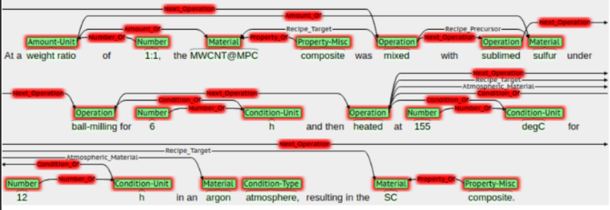
Grounded Action Graph (K):

* An action sequence K is defined to be a sequence of valid actions (or production rules) rooted at some initial state s_0:

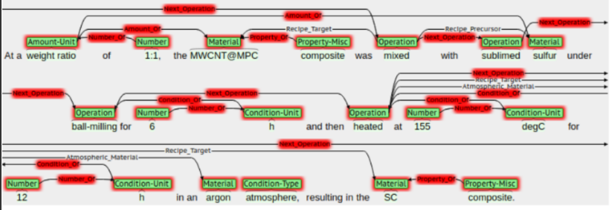
$$K = (s_0, \lambda_0, \lambda_1, ..., \lambda_n) \tag{1}$$

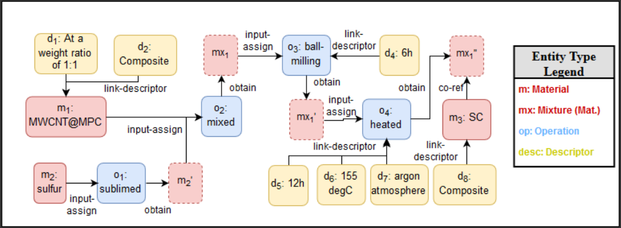
>>> What's in a story?

>>> What's in a story?

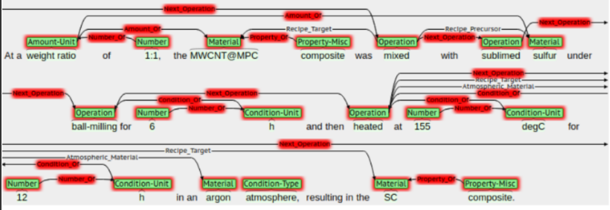


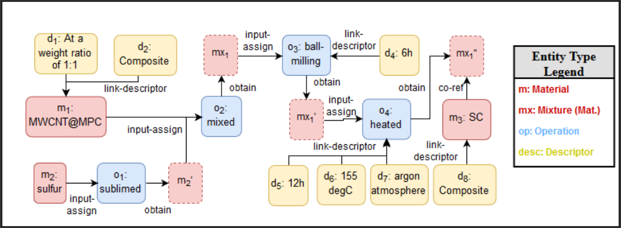
>>> What's in a story?





>>> What's in a story?






```
s_0 = { at(m1, r), at(m2, r), ... }

1. link-descriptor(d1, m1)
2. link-descriptor(d2, m1)
3. input-assign(m1, o2)
4. input-assign(m2, o1)
5. run-op(o1)
6. obtain-output(o1)
7. input-assign(m2, o2)
8. run-op(o2)
9. obtain-output(o2)
10. input-assign(mx1, o3)
11. link-descriptor(d4, o3)
12. run-op(o3)
13. obtain-output(o3)
14. input-assign(mx1, o4)
15. link-descriptor(d5, o4)
16. link-descriptor(d6, o4)
17. link-descriptor(d7, o4)
18. run-op(o4)
19. obtain-output(o4)
20. co-ref(m3, mx1)
21. link-descriptor(d8, m3)
```

>>> Learning Task

Narrative Comprehension Objective:

>>> Learning Task

Narrative Comprehension Objective:

* Learn a mapping $X \rightarrow K$

>>> Learning Task

Narrative Comprehension Objective:

* Learn a mapping $X \rightarrow K$
* May be highly complex mapping!

>>> Learning Task

Narrative Comprehension Objective:

* Learn a mapping $X \rightarrow K$
* May be highly complex mapping!
* Approach: Structured prediction. Map input X to an intermediate enriched text-based game representation G, interpret X as its instructions, K as its solution!

>>> Learning Task

Narrative Comprehension as Game Solving:

>>> Learning Task

Narrative Comprehension as Game Solving:

* From an input surface X, infer a game G and its initial state s_0.

$$X \to (G, s_0) \to K$$

>>> Learning Task

Narrative Comprehension as Game Solving:

* From an input surface X, infer a game G and its initial state s_0.

$$X \to (G, s_0) \to K$$

* G is a text-based game, modelled as a POMDP: $G = (S, A, T, \Omega, O, R, \gamma)$.

>>> Learning Task

Narrative Comprehension as Game Solving:

* From an input surface X, infer a game G and its initial state s_0.

$$X \to (G, s_0) \to K$$

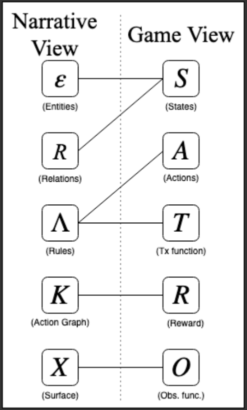
* G is a text-based game, modelled as a POMDP: $G = (S, A, T, \Omega, O, R, \gamma)$.
* The sequence of actions solving G will be exactly the desired K!

>>> Learning Task

Narrative Comprehension as Game Solving:

* From an input surface X, infer a game G and its initial state s_0.

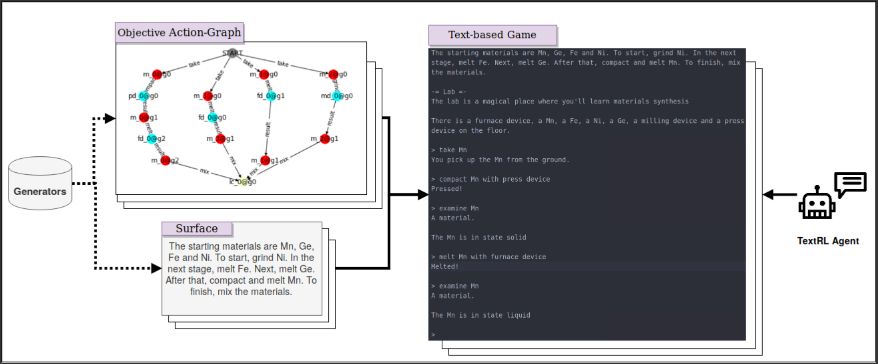
$$X \to (G, s_0) \to K$$



* G is a text-based game, modelled as a POMDP: $G = (S, A, T, \Omega, O, R, \gamma)$.
* The sequence of actions solving G will be exactly the desired K!

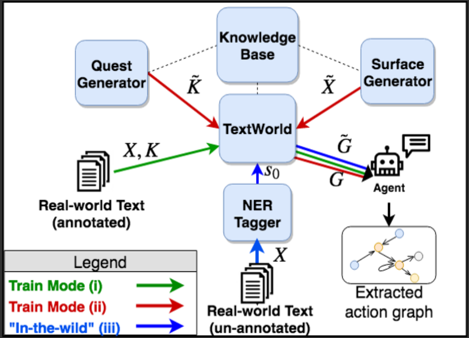
>>> Solution Architecture

* Generate quests using classic AI planning & search.
* Structured Natural Language Generation (NLG) for generating surfaces corresponding to quests.
* Microsoft's TextWorld[1] sandbox environment for generating text-based games for training RL agents.



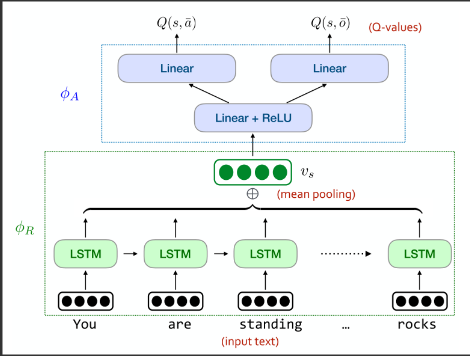
>>> Solution Architecture

* Generate quests using classic AI planning & search.
* Structured Natural Language Generation (NLG) for generating surfaces corresponding to quests.
* Microsoft's TextWorld[1] sandbox environment for generating text-based games for training RL agents.

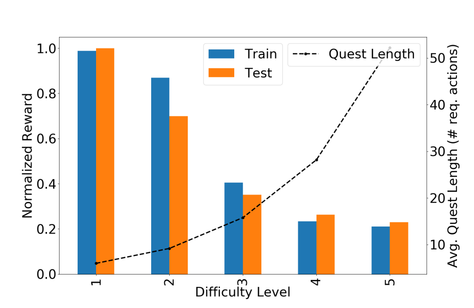


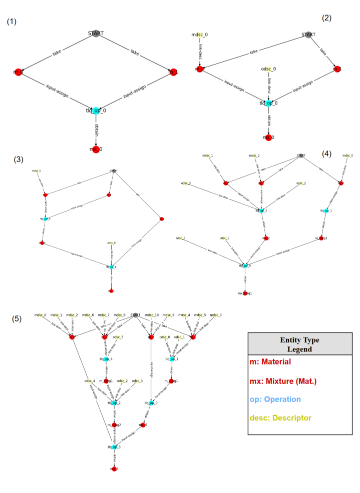
>>> Preliminary Experiments

* Applied basic text-based reinforcement learning agent on generated materials synthesis "quests".
* Based on LSTM-DQN architecture from [3]
* "Friendly environment" can provide action pruning, (in future) state tracking, curriculum learning.



>>> Results


Train
Test
Quest Length
Normalized Reward
Avg. Quest Length (# req. actions)
Difficulty Level
0.0
0.2
0.4
0.6
0.8
1.0
10
20
30
40
50
1
2
3
4
5


(1)
(2)
(3)
(4)
(5)
START
take
input-assign
obtain
link-desc
mdsc_0
odsc_0
Entity Type Legend
m: Material
mx: Mixture (Mat.)
op: Operation
desc: Descriptor

>>> Conclusions

>>> Conclusions

* Growing recognition of need for symbolic reasoning capabilities for complex reading comprehension tasks.

>>> Conclusions

* Growing recognition of need for symbolic reasoning capabilities for complex reading comprehension tasks.
* In procedural text understanding, key obstacles to work towards this are lack of suitable training environments and fully annotated training data.

>>> Conclusions

* Growing recognition of need for symbolic reasoning capabilities for complex reading comprehension tasks.
* In procedural text understanding, key obstacles to work towards this are lack of suitable training environments and fully annotated training data.
* Neural programming, reinforcement learning and structured NLG all fast growing fields and our system can benefit directly from this.

>>> Conclusions

* Growing recognition of need for symbolic reasoning capabilities for complex reading comprehension tasks.
* In procedural text understanding, key obstacles to work towards this are lack of suitable training environments and fully annotated training data.
* Neural programming, reinforcement learning and structured NLG all fast growing fields and our system can benefit directly from this.
* While niche application, setting has much in common with fundamental challenges in natural language understanding $\rightarrow$ still long way to go...

>>> Sample Training Games

* After 2 epochs
* After 40 epochs

\>>> Check out our paper for more details!

Playing by the Book: An Interactive Game Approach for Action Graph Extraction from Text

Ronen Tamari, Hiroyuki Shindo, Dafna Shahaf, Yuji Matsumoto

https://arxiv.org/abs/1811.04319

>>> References I

- M.-A. Côté, A. Kádár, X. Yuan, B. Kybartas, T. Barnes, E. Fine, J. Moore, M. Hausknecht, L. E. Asri, M. Adada, W. Tay, and A. Trischler.
  Textworld: A learning environment for text-based games.
  CoRR, abs/1806.11532, 2018.
- R. W. Langacker and G. Lakoff.
  Women, Fire, and Dangerous Things: What Categories Reveal about the Mind, volume 64.
  1988.
- K. Narasimhan, T. Kulkarni, and R. Barzilay.
  Language Understanding for Text-based Games Using Deep Reinforcement Learning.
  jun 2015.

>>> References II

- F. Scarselli, M. Gori, A. C. Tsoi, M. Hagenbuchner, and G. Monfardini.
  The graph neural network model.
  IEEE Transactions on Neural Networks, 20:61--80, 2009.
- B. D. Trisedya, J. Qi, R. Zhang, and W. Wang.
  GTR-LSTM: A Triple Encoder for Sentence Generation from RDF Data.
  Proceedings of ACL, pages 1627--1637, 2018.
- Y. Yao, J. Xu, J. Shi, and B. Xu.
  Learning to activate logic rules for textual reasoning.
  Neural Networks, 106:42 -- 49, 2018.

>>> Structured NLG

RDF triples	⟨John Doe, birth place, London⟩ ⟨John Doe, birth date, 1967-01-10⟩ ⟨London, capital of, England⟩
Target sentence	John Doe was born on 1967-01-10 in London, the capital of England.

Table 1: RDF based sentence generation.

Figure: Sample structured text generation from [5]

⇐