Acquiring and Modeling
Abstract Commonsense Knowledge via Conceptualization

Yangqiu Song
Department of CSE, HKUST

Work done with Mutian He, Tianqing Fang, Weiqi Wang.
Outline

• Motivation

• Abstract ATOMIC
“Commonsense Knowledge”

• When we communicate,
  • we omit a lot of “common sense” knowledge, which we assume the hearer/reader possesses
  • we keep a lot of ambiguities, which we assume the hearer/reader knows how to resolve

• Attributes of objects
  • A lemon is sour.

• Condition/consequence of actions
  • To open a door, you must usually first turn the doorknob.

• Cause/effect between events and states
  • If you forget someone’s birthday, they may be unhappy with you.

• Social:
  • If you forget your friend’s birthday, he/she may be mad at you.

• Physical, temporal, spatial:
  • Apples fall instead of floating in the air.

• World entities:
  • Lions are bigger than cats.
Many Possible Applications

• How to plan a wedding, what to do and what to buy?
  • Understand the timeline
  • Understand the events
    • Precedents
    • Consequences
  • Understand the people
    • Social conventions

18 MONTHS to go...
- Get the inspiration book "Your Day Your Way" by Rock My Wedding.
- Decide how to centre your inspiration.
  Maybe a folder, Pinterest, Instagram.
- Work out a budget.
- Work out guest list and choose bridal party.

16 MONTHS to go...
- Look at venues and check availability.
- Book church, hall, etc. and sort wedding licence.
- Start researching suppliers via Rock My Wedding recommended suppliers.
- Sort wedding insurance.

14 MONTHS to go...
- Arrange appointments at wedding dress boutiques.
- Book a wedding planner if you want one and budget allows.
- Book your photographer and videographer.

12 MONTHS to go...
- Consider underwear and try on dresses.
- Look at honeymoon options.

9 MONTHS to go...
- Book florist.
- Send Save The Dates.
- Look into Groom's attire.

6 MONTHS to go...
- Shop for Bridalmaid dresses.
- Confirm catering.
- Taste cakes and book.
- Book entertainment: musical and other.
- Book stationery with a professional or plan your DIY stationery.
- Consider transport.
- Research and book any items you may need to hire.
- Decide on a gift list company and register.
- Book hair and make-up trials.

4 MONTHS to go...
- Book your Underwear if you didn't get it before trying on dresses.
- Shop for shoes.
- Buy Groom's suit and leave time for alterations.
- Send invitations.
- Choose wedding rings.
- Have hair and make-up trials and book.

3 WEEKS to go...
- Arrange your seating plan.
- Start making your table plan if you're making yourself.
- Write vows.
- Book beauty and spa treatments.
- Collect wedding rings.

1 WEEK to go...
- Arrange your seating plan.
- Start making your table plan if you're making yourself.
- Write vows.
- Pack for honeymoon.
- Book beauty and spa treatments.
- Collect wedding outfits.

THE DAY BEFORE
- Drop off any decor items to the venue.
- Go through roles with everyone.
- Have a good meal and make sure you have arranged breakfast for the morning.
- Get an early night and have an amazing day!
How to Define Commonsense Knowledge as Computer Scientists?

• According to Liu & Singh (2004)

• “While to the average person the term ‘commonsense’ is regarded as synonymous with ‘good judgement’, ”

• “the AI community it is used in a technical sense to refer to the millions of basic facts and understandings possessed by most people.”

H Liu and P Singh, ConceptNet - a practical commonsense reasoning tool-kit, BTTJ, 2004
ConceptNet: An Approach Developed 18 Years Ago

- ConceptNet5 (Speer and Havasi, 2012)
- Core is from Open Mind Common Sense (OMCS) (Liu & Singh, 2004)

Essentially a crowdsourcing based approach + text mining

H Liu and P Singh, ConceptNet - a practical commonsense reasoning tool-kit, BTTJ, 2004
ATOMIC: Everyday If-then Commonsense Knowledge

- **Crowdsourcing** 9 Types of IF-THEN inferential knowledge
- All personal entity information has been removed to reduce ambiguity
- Mostly arbitrary texts

Usually in the Form of a Knowledge Graph

• Commonsense knowledge graph (CKG): represented in triples of texts
  • $\langle h: \text{PersonX is hungry}, r: \text{xWant (then PersonX wants)}, t: \text{to have lunch} \rangle$

• An excerpt of the world with prototypical events and causes/consequences
  • Correspond to real-world situations

• But how could we know if it is generalizable?
Limited Coverage: Symbolic CKG

- CKGs can’t cover all the entities and situations
  - Not to say corresponding triples

Reasonable consequences for “PersonX gets the flu” are not covered by “PersonX gets a cold”

Source: https://mosaickg.apps.allenai.org/kg_atomic2020/
Conceptualization: A Missed Point

• Countless Entities and Situations in the real world
  • So many different things and situations we encounter new things everyday

• Humans understand the world through concepts
  • Summarize previous experiences into abstract mental representation
    • Entity concepts: animal (cat, dog, pet, tiger, …)
    • Situational concepts: a relaxing event (have a cup of coffee, take a break, …)

• CKGs are not enough (checked with Probase)

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>#Head</td>
<td>24.3K</td>
<td>44.0K</td>
<td>1,103.0K</td>
</tr>
<tr>
<td>#Triple</td>
<td>793.3K</td>
<td>1246.6K</td>
<td>3,235.9K</td>
</tr>
<tr>
<td>Average Degree</td>
<td>32.6</td>
<td>28.4</td>
<td>2.9</td>
</tr>
<tr>
<td>Concept Coverage</td>
<td>0.34%</td>
<td>0.76%</td>
<td>8.00%</td>
</tr>
<tr>
<td>Average Distinct Concept</td>
<td>0.093</td>
<td>0.114</td>
<td>0.048</td>
</tr>
</tbody>
</table>

Data are available at [https://concept.research.microsoft.com/](https://concept.research.microsoft.com/)
Limited Coverage: Neural Models

• Neural commonsense models to handle arbitrary texts?

COMET fails to predict a common cause for cold and chickenpox (“to be around someone sick”), even though both are instances of contagious disease.

Common consequence of skin disease (“get a rash”) fails to generalize to chickenpox.

Superficial textual inferences like “PersonX eats chicken” is included.
Commonsense Reasoning in General

• Conceptual induction
  • Conceptualization and compositionality are keys to commonsense reasoning (generalization), but there is still lack of study

X: Item does not fit in container, REASON, item is big
  Trophy is an item; Suitcase is a container

Y: Trophy does not fit in suitcase, REASON, it is big

Current deep learning usually just preforms induction by learning from examples, and then performs verification of a new claim, instead of decuction.

CSKB/Training Data
  • Computer not fit in parcel, REASON, Computer is big
  • Rock not fit in carrier, REASON, rock is big
  • ...

• If we can explicitly instantiate new claims from abstractive claims, then we can have much more examples for learning models to learn.
• Conceptual induct is yet to be too difficult for language models: so many concepts
Conceptualization: Related Theories

• Vagueness
  • No strict hierarchy
  • Effort-taking on borderline cases
  • Diverse and context-dependent

• Abstraction of the world
  • K-line theory: not only conceptualize entities, but also events and mental states

Representing Knowledge in Multiple Ways

**SituationNet:**

detailed descriptions of situations

Prototypical situations

---


Figure Credit: [https://ocw.mit.edu/courses/media-arts-and-sciences/mas-961-ambient-intelligence-spring-2005/lecture-notes/week4_push_singh.pdf](https://ocw.mit.edu/courses/media-arts-and-sciences/mas-961-ambient-intelligence-spring-2005/lecture-notes/week4_push_singh.pdf)
Representing Knowledge in Multiple Ways

- Encode memories in "abstract" form.
- Search all memory for the "nearest match."
- Remember "methods," not "answers."

The K-Line Theory

• Attach a K-node (a mental state, KE) to a “Pyramid” agent (PE) at a certain level
  • The pyramid is a tree structure that we use to conceptualize the world
  • The mapping has a lower-band limit and a higher-band limit, to compare the right common, non-conflicting properties
• Then the PE will help us to make abstraction, logical and procedural reasoning

When comparing Tesla with Google, Toyota, some small company, we need the right level and right perspective of comparison
• E.g., mapping Tesla to a company, big company, IT company, AI company, high-tech company, automobile company

Outline

• Motivation

• Abstract ATOMIC: to develop a commonsense knowledge graph with more abstraction and conceptual induction capability
Abstract Events

• We call the original triple in ATOMIC the **situational knowledge**
  • \( \langle h: \text{PersonX is hungry}, r: x\text{Want} \text{ (then PersonX wants)}, t: \text{to have lunch} \rangle \)
  • Head is usually a description regarding everyday events on some unspecified PersonX
  • Tails are less complete as sentences and more difficult to parse for conceptualization

• Event conceptualization:
  • A **conceptualization** is an abstract simplified view of some selected part of the world
  • **Different levels of abstractness**
    • “PersonX drinks coca cola” as “[drinking coca cola],” “[drinking beverage],” “[event]”
  • **Different perspectives**
    • “Coca cola” as “[sugary beverage],” “[phosphate containing beverage],” “[iced drink],” not in a strict hierarchical taxonomy
      • PersonX drinks [iced drink], xReact, refreshed
      • PersonX drinks [sugary beverage], xEffect, gain weight
Abstract Events

• Formally defined as a textual template with a slot, filled by a concept, like “PersonX drinks [beverage],” “[drinking beverage],” “[relaxing event]”

• We construct a “bipartite graph” between situational knowledge and abstract knowledge
Abstract Triples

- Triples with an Abstract Event as the Head
  - E.g., link “PersonX drinks [sugary beverage]” to an effect of “gain weight”
Validity

• Situational events and triples
  • Defeasibility of commonsense: each piece of commonsense knowledge is often termed as factoid that is only plausible, or typically true, according to human intuition without more formal reasoning
    • “a dog is smaller than a person” could be a false claim but is a commonsense

• Abstract events and triples
  • By whether it is typically valid among instantiations
    • PersonX has a cup of coffee \(\rightarrow\) [relaxing event]
  • Even if itself is not ordinary CKG event or triple in natural language
    • PersonX drinks [phosphate containing beverage]
    • PersonX spends [time interval] reading

• Event conceptualizations
  • By whether it covers the meaning in the context
    • PersonX has a cup of coffee \(\rightarrow\) PersonX has [cup]?
Overall Running Example

Situational Knowledge

- **PersonX takes a short break**
  - xWant to have more

- **PersonX sees a funny video**
  - xWant to share it

- **PersonX drinks coca cola**
  - xIntent to quench thirst

- **PersonX has a cup of tea**
  - xNeed to boil water

- **PersonX drinks red bull**
  - xReact to energized

- **PersonX drinks coca cola**
  - xEffect to gain weight

1. **Identification**
Overall Running Example
Concept Bank

• Use both WordNet and Probase
  • To cover the flexibility of human conceptualization

Data are available at https://concept.research.microsoft.com/
Data are available at https://concept.research.microsoft.com/


Slide Credit: Haixun Wang
Outline

• Motivation

• Abstract ATOMIC
  • Construction steps
Step 1: Identification

• Identify candidates
  • Word matching is not enough
    • “She gives her pet food” (She gives food to her pet)

• Work on dependency tree
  • Check each constituent/subtree
  • Nominal candidate or predicative candidate
  • Determined by linguistic tags by a set of rules (customized to ATOMIC data)
    • E.g., if a constituent is a noun according to POS, and a direct object of a verb, then it should be an entity
Step 2: Conceptualization (I: Concept Linking)

- To link an entity or situation to the concept bank
  - Require sophisticated natural language understanding

- Possibly distinct from the text
  - “a cup of coffee” -> coffee (instead of the head word cup)
  - “Alice lives with her boyfriend” -> cohabitation (instead of simply living)
Step 2: Conceptualization (I: Heuristic Rules)

- Rules to derive possible concepts
  - As WordNet or Probase nodes
  - Use linguistic features for the constituent
  - Aided by WordNet relations, NOMBANK, and GlossBERT

- Impossible to be accurate
  - Goal: to provide good candidates

<table>
<thead>
<tr>
<th>Type</th>
<th>Example</th>
<th>Concepts</th>
<th>Method</th>
</tr>
</thead>
<tbody>
<tr>
<td>Word</td>
<td>PersonX finds some cats</td>
<td>cat</td>
<td>Directly use the headword, possibly lemmatized</td>
</tr>
<tr>
<td>Compound/Phrase</td>
<td>PersonX sees many stray cats</td>
<td>stray cat</td>
<td>Collect compounds or phrases in nominal candidates by word matching in the constituent, subject to inflections</td>
</tr>
<tr>
<td>Predicate (Verb)</td>
<td>PersonX drinks coffee</td>
<td>drinking</td>
<td>Directly use the gerund of the verb</td>
</tr>
<tr>
<td></td>
<td>PersonX says he enjoys himself</td>
<td>enjoyment</td>
<td>Check WordNet and NOMBANK for the noun form of the verb</td>
</tr>
<tr>
<td>Predicate (Adj.)</td>
<td>PersonX is happy</td>
<td>happiness</td>
<td>Same as above, for a copula with adjective complement</td>
</tr>
<tr>
<td>Conjunction</td>
<td>PersonX sees doctors and nurses</td>
<td>doctor, nurse</td>
<td>Use concepts from both conjuncts</td>
</tr>
<tr>
<td>Nominal Classifier</td>
<td>PersonX has a cup of tea</td>
<td>tea</td>
<td>Directly return results from the accompanied argument, if the head and the preposition form a NOMBANK transparent construction</td>
</tr>
<tr>
<td></td>
<td>PersonX sees a group of people</td>
<td>group, people</td>
<td>If an argument is connected to another one by “of” but not a NOMBANK transparent construction, both the head and accompanied argument are used.</td>
</tr>
<tr>
<td>Verb Phrase</td>
<td>PersonX drinks coffee</td>
<td>drinking</td>
<td>Combine the verb with its arguments</td>
</tr>
<tr>
<td>Phrasal Verb</td>
<td>PersonX gets up late</td>
<td>get up</td>
<td>Check WordNet for combination of the verb and one or two particle in the text</td>
</tr>
<tr>
<td>Light Verb</td>
<td>PersonX gives a speech</td>
<td>giving, speech</td>
<td>For light verbs like give, take, have, etc., the predicand can be the actual concept</td>
</tr>
<tr>
<td></td>
<td>PersonX goes shopping</td>
<td>shopping</td>
<td>Directly use results from the predicand when it is a ground</td>
</tr>
<tr>
<td>Raising-to-subject</td>
<td>PersonX seems to be happy</td>
<td>happiness</td>
<td>Directly use results from the predicand for cases like seem, appear, used, etc.</td>
</tr>
<tr>
<td>Control Verb</td>
<td>PersonX wants to leave</td>
<td>want, leave</td>
<td>Use both the head (superordinate verb) and results from its complement</td>
</tr>
<tr>
<td>Adj. + Infinitive</td>
<td>PersonX is likely to leave</td>
<td>leave</td>
<td>Directly use results from the infinitive for a copula with some adjectives like likely, going, about, able, etc., as complement</td>
</tr>
</tbody>
</table>

Annotation Phase 1

PersonX finds a stray cat

1. What concept does “a stray cat” describe?
   - stray cat

2. Which ones are plausible abstractions for the given concept in the context?
   - stray cat
     - predator
     - specie
     - pest
     - animal problem
     - mammalian nest predator
     - cat

Only concepts found in Probase will be allowed to be submitted
Annotation Phase 2

PersonX finds a stray cat

Do the sentences below cover the meaning above?

PersonX finds stray cat
PersonX finds animal
PersonX finds animal problem
PersonX finds free living animal
PersonX finds critter
PersonX finds free-living animal
PersonX finds mammalian nest predator
PersonX finds predator
PersonX finds wild animal
PersonX finds feline
PersonX finds specie
PersonX finds mammalian

Feel free to add these determiners before the substitution to make it valid: the, a, PersonXs, some, the event of, the action of, at (the state of), with (the attribute of), ...

Submit
Annotation Phase 2

I heard that PersonX gets hit by a car

Do the sentences below cover the meaning above?

- I heard about (the/PersonX's) accident  [Yes/No]
- I heard about (the/PersonX's) emergency  [Yes/No]
- I heard about (the/PersonX's) traffic incident  [Yes/No]
- I heard about (the/PersonX's) offense  [Yes/No]
- I heard about (the/PersonX's) serious violation  [Yes/No]
- I heard about (the/PersonX's) traffic offense  [Yes/No]
- I heard about (the/PersonX's) emergency situation  [Yes/No]
- I heard about (the/PersonX's) unexpected event  [Yes/No]
- I heard about (the/PersonX's) car accident  [Yes/No]
- I heard about (the/PersonX's) event  [Yes/No]
- I heard about (the/PersonX's) injury  [Yes/No]
- I heard about (the/PersonX's) traumatic event  [Yes/No]

Feel free to add these determiners before the substitution to make it valid: the, a, PersonX's, some, the event of, the action of, at (the state of), with (the attribute of), ...
Annotation: Quality Control

• Rigorous worker enrollment
  • 95% acceptance on at least 1000 tasks
  • One or two qualification tests

• Detailed instructions
  • Over 30 examples

• In-progress monitoring
  • Disqualify underperformed workers, and discard their annotations
Annotation Results

• For 24.3K events in ATOMIC, after filtering general concepts, idioms, and duplicates, 15.9K events are valid candidates.

• Then 10K events (with entities or situations to be conceptualized) are randomly selected, from 8,045 original ATOMIC events (around 1/3 of ATOMIC heads).

• After annotation, 7,019 ATOMIC events were used to form 18,964 different positive abstract events.

<table>
<thead>
<tr>
<th></th>
<th>Event</th>
<th>Triple</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total Questions</td>
<td>131,004</td>
<td>90,207</td>
</tr>
<tr>
<td>Questions with Agreement</td>
<td>92,235</td>
<td>81,197</td>
</tr>
<tr>
<td>Positives</td>
<td>40,833</td>
<td>65,900</td>
</tr>
<tr>
<td>Positive Rate</td>
<td>44.27%</td>
<td>81.16%</td>
</tr>
<tr>
<td>Inter-annotator Agreement</td>
<td>71.4%</td>
<td>73.0%</td>
</tr>
<tr>
<td>Manual Inspection Agreement</td>
<td>87.0%</td>
<td>86.5%</td>
</tr>
</tbody>
</table>
Conceptualization: Rules and Models

• Conceptualization needs both NLU and taxonomic knowledge
• Rules use taxonomic KG explicitly
  • But lack of contexts
• Neural models are contextualized
  • Doubtful diversity
  • ...even the training data is aided by the taxonomic KG
• We combine both approaches and introduce a gatekeeper module
Step 2: Conceptualization (II: Neural Concept Generator)

• Use prompt to directly generate possible abstractions
  • With a neural autoregressive generator
    • PersonX buys <c> a hot dog </c>. <c> hot dog </c> is an instance of [GEN]
  • Trained by our data
    • GPT2-base, 32 batch size, learning rate 1e-5
    • Measured by BLEU-2 from 10-beam search

(a) Architecture of LM generator for Event Conceptualization based on pretrained GPT-2.
Neural Concept Generator: Results

• Trained on the annotated data
  • Split by the original ATOMIC event partition
  • Use best models on its dev set
  • Alternative prompts and hyper-parameters show no improvements

<table>
<thead>
<tr>
<th>BLEU-1</th>
<th>Dev Set</th>
<th>Test Set</th>
</tr>
</thead>
<tbody>
<tr>
<td>Supervised Generator</td>
<td>65.1</td>
<td>68.0</td>
</tr>
<tr>
<td>GPT2 Zero-shot</td>
<td>25.0</td>
<td>20.4</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>BLEU-2</th>
<th>Dev Set</th>
<th>Test set</th>
</tr>
</thead>
<tbody>
<tr>
<td>Supervised Generator</td>
<td>61.1</td>
<td>56.5</td>
</tr>
<tr>
<td>GPT2 Zero-shot</td>
<td>4.8</td>
<td>2.6</td>
</tr>
</tbody>
</table>
Step 2: Conceptualization (III: Verifier)

- Gatekeeping all event abstractions we found
  - With our annotated data
  - RoBERTa-base, 64 batch size, learning rate 2e-5
  - Measured by accuracy, threshold from dev set
Verifier: Results

• Baselines
  • Negative Sampling
    • Using positive samples only, 5 negatives per positive
  • Generator for discrimination
    • Use losses as scores

<table>
<thead>
<tr>
<th>Accuracy</th>
<th>Dev Set</th>
<th>Test Set</th>
</tr>
</thead>
<tbody>
<tr>
<td>Supervised Learning</td>
<td>83.9</td>
<td>85.0</td>
</tr>
<tr>
<td>Negative Sampling</td>
<td>78.4</td>
<td>78.8</td>
</tr>
<tr>
<td>Generator – Supervised (GPT-2)</td>
<td>76.9</td>
<td>77.6</td>
</tr>
<tr>
<td>Generator - Zero-shot (GPT-2)</td>
<td>55.1</td>
<td>58.5</td>
</tr>
</tbody>
</table>
## Examples

<table>
<thead>
<tr>
<th>Event</th>
<th>Linked concepts</th>
<th>Conceptualization (Heuristic linking)</th>
<th>Conceptualization (Neural Generation)</th>
</tr>
</thead>
<tbody>
<tr>
<td>PersonX buys [a new iphone]</td>
<td>iphone, new iphone</td>
<td>IOS device, smart phone, mobile devise, device, apple device, product, ...</td>
<td>smart phone, phone, product, electronic device, device, mobile device</td>
</tr>
<tr>
<td>PersonX pays [PersonX’s water bill]</td>
<td>bill, water bill</td>
<td>bill, expense, basic household expense, user charge, utility bill, ...</td>
<td>bill, expense, ...</td>
</tr>
<tr>
<td>[PersonX gets a cold]</td>
<td>contracting, cold</td>
<td>cold, common illness, illness, infection, minor ailment, respiratory infection, upper respiratory infection, viral infection, ...</td>
<td>cold, condition, sickness, ...</td>
</tr>
<tr>
<td>[PersonX gives birth to children]</td>
<td>giving birth, birth, gift</td>
<td>occasion, birth, life change, life event, happy event, ...</td>
<td>creation, birth, event, life event, life cycle event, life changing event</td>
</tr>
<tr>
<td>[PersonX has a bad day at work]</td>
<td>experience</td>
<td>experience</td>
<td>difficulty, negative event, problem, unpleasant experience, ...</td>
</tr>
</tbody>
</table>
Step 3

Inference
Verifier

PersonX drinks a cup of tea

Heuristic Identification

Conceptualization

Neural Concept Generator

Link

Generate

Taxonomy

Abstract Concept

PersonX drinks [caffeinated drink]

Conceptualization Verifier

PersonX drinks [a cup of tea] [SEP] caffeinated drink

PersonX drinks [a cup of tea] [SEP] crop

Event Conceptualizations

Abstract Event

Abstract Triple

xIntent
get refreshed
xNeed
find a cup

Discriminate
Inference

• Collect all tails from the instantiations in ATOMIC

• Verify if it applies to the abstract event
  • With a neural model on our annotated data
Question 1

If PersonX asks PersonY [marriage], after that/as result, the Effect on PersonX is: personX puts a ring on personY's finger.

How likely is this sentence going to happen? (Invalid if it does not make sense to you)

- Always / Usually
- Typical: Often / Probable
- Atypical: Farfetched / Never
- Invalid: Can't understand the meaning
Annotation Results

- Based on the positive abstract events, total 1,149,118 possible abstract triples are collected
- 2,756 abstract events are sampled for annotation

<table>
<thead>
<tr>
<th>Event</th>
<th>Event</th>
<th>Triple</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total Questions</td>
<td>131,004</td>
<td>90,207</td>
</tr>
<tr>
<td>Questions with Agreement</td>
<td>92,235</td>
<td>81,197</td>
</tr>
<tr>
<td>Positives</td>
<td>40,833</td>
<td>65,900</td>
</tr>
<tr>
<td>Positive Rate</td>
<td>44.27%</td>
<td>81.16%</td>
</tr>
<tr>
<td>Inter-annotator Agreement</td>
<td>71.4%</td>
<td>73.0%</td>
</tr>
<tr>
<td>Manual Inspection Agreement</td>
<td>87.0%</td>
<td>86.5%</td>
</tr>
</tbody>
</table>
Inference Verification Modeling

• Similar to conceptualization verifier
• Adopt RoBERTa with following prompts

<table>
<thead>
<tr>
<th>Relation</th>
<th>Prompt</th>
</tr>
</thead>
<tbody>
<tr>
<td>xNeed</td>
<td>Before that PersonX needs:</td>
</tr>
<tr>
<td>xIntent</td>
<td>PersonX’s intention is:</td>
</tr>
<tr>
<td>xAttr</td>
<td>PersonX will be described as:</td>
</tr>
<tr>
<td>xEffect</td>
<td>Effects on PersonX will be:</td>
</tr>
<tr>
<td>xWant</td>
<td>After that PersonX wants:</td>
</tr>
<tr>
<td>xReact</td>
<td>After that PersonX feels:</td>
</tr>
<tr>
<td>oEffect</td>
<td>Effects on others will be:</td>
</tr>
<tr>
<td>oWant</td>
<td>After that others want:</td>
</tr>
<tr>
<td>oReact</td>
<td>After that others feel:</td>
</tr>
</tbody>
</table>

• Measured by AUC
• Other prompts and hyperparameters attempted as well
Inference Verification Modeling: Results

• Baselines
  • Negative Sampling
    • Positive examples from ATOMIC (same size as annotated data)
    • Positive examples from our annotated positives
    • Mixed
  • ATOMIC generator: COMET (GPT2-medium), 30.3 BLEU-2 on ATOMIC subset

<table>
<thead>
<tr>
<th>AUC</th>
<th>Dev Set</th>
<th>Test Set</th>
</tr>
</thead>
<tbody>
<tr>
<td>ATOMIC + Negative Sampling</td>
<td>0.62</td>
<td>0.65</td>
</tr>
<tr>
<td>Annotation</td>
<td>0.72</td>
<td>0.74</td>
</tr>
<tr>
<td>w/ Negative Sampling</td>
<td>0.67</td>
<td>0.67</td>
</tr>
<tr>
<td>Annotation + ATOMIC</td>
<td>0.73</td>
<td>0.76</td>
</tr>
<tr>
<td>w/ Negative Sampling</td>
<td>0.63</td>
<td>0.65</td>
</tr>
<tr>
<td>Generator based on ATOMIC</td>
<td>0.49</td>
<td>0.50</td>
</tr>
</tbody>
</table>

Another experiment on knowledge base population shows generative model is as effective as KB completion models.

Supervised Learning

<table>
<thead>
<tr>
<th>Model</th>
<th>Dev Set</th>
<th>Test Set</th>
</tr>
</thead>
<tbody>
<tr>
<td>KG-BERT (BERT-base) 110M</td>
<td>62.5</td>
<td></td>
</tr>
<tr>
<td>KG-BERT (BERT-large) 340M</td>
<td>67.7</td>
<td></td>
</tr>
<tr>
<td>KG-BERT (DeBERTa-base) 10M</td>
<td>64.5</td>
<td></td>
</tr>
<tr>
<td>KG-BERT (DeBERTa-large) 350M</td>
<td>69.2</td>
<td></td>
</tr>
<tr>
<td>KG-BERT (BART-base) 139M</td>
<td>65.1</td>
<td></td>
</tr>
<tr>
<td>KG-BERT (BART-large) 406M</td>
<td>70.4</td>
<td></td>
</tr>
<tr>
<td>KG-BERT (RoBERTa-base) 110M</td>
<td>68.0</td>
<td></td>
</tr>
<tr>
<td>KG-BERT (RoBERTa-large) 340M</td>
<td>70.9</td>
<td></td>
</tr>
<tr>
<td>COMET (GPT2-small) 117M</td>
<td>69.6</td>
<td></td>
</tr>
<tr>
<td>COMET (GPT2-medium) 345M</td>
<td>69.7</td>
<td></td>
</tr>
<tr>
<td>COMET (GPT2-large) 774M</td>
<td>70.6</td>
<td></td>
</tr>
<tr>
<td>COMET (GPT2-XL) 1558M</td>
<td>70.7</td>
<td></td>
</tr>
</tbody>
</table>
Abstract ATOMIC

• Selection scores from neural models
  • Event conceptualization score > 0.8
  • Triple score > 0.9

Heuristic concept linker produce much more diverse candidates but much less accurate

<table>
<thead>
<tr>
<th>Numbers of selected data</th>
<th>0.7~0.8</th>
<th>0.8~0.9</th>
<th>≥0.9</th>
</tr>
</thead>
<tbody>
<tr>
<td>Event Conceptualization (by Neural Concept Generator)</td>
<td>10.3K</td>
<td>17.7K</td>
<td>171.1K</td>
</tr>
<tr>
<td>Event Conceptualization (by Heuristic Concept Linker)</td>
<td>8.3K</td>
<td>11.5K</td>
<td>81.3K</td>
</tr>
<tr>
<td>Event Conceptualization (Total)</td>
<td>16.7K</td>
<td>26.2K</td>
<td>203.0K</td>
</tr>
<tr>
<td>Different Abstract Event</td>
<td>4.3K</td>
<td>7.0K</td>
<td>63.0K</td>
</tr>
<tr>
<td>Abstract Triple</td>
<td>542.2K</td>
<td>937.2K</td>
<td>2,947.9K</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Human evaluation based on sampled data</th>
<th>0.7~0.8</th>
<th>0.8~0.9</th>
<th>≥0.9</th>
</tr>
</thead>
<tbody>
<tr>
<td>Event Conceptualization (by Neural Concept Generator)</td>
<td>0.64</td>
<td>0.72</td>
<td>0.88</td>
</tr>
<tr>
<td>Event Conceptualization (by Heuristic Concept Linker)</td>
<td>0.67</td>
<td>0.74</td>
<td>0.90</td>
</tr>
<tr>
<td>Abstract Triple</td>
<td>0.41</td>
<td>0.55</td>
<td>0.71</td>
</tr>
</tbody>
</table>

ATOMIC’s human score: 86.18%
<table>
<thead>
<tr>
<th>Event</th>
<th>Instantiation</th>
<th>Relation</th>
<th>Positive Tails</th>
<th>Negative Tails</th>
</tr>
</thead>
<tbody>
<tr>
<td>PersonX calls [health professional]</td>
<td>the doctor, the dentist</td>
<td>xWant</td>
<td>set an appointment, to ask the doctor a question, to tell the doctor their problems, ...</td>
<td>to take their pet there, to ask a question</td>
</tr>
<tr>
<td></td>
<td></td>
<td>xIntent</td>
<td>to schedule an appointment, to help pet, to be healthy, to feel better, ...</td>
<td>to know about their pet, to be informed</td>
</tr>
<tr>
<td></td>
<td></td>
<td>xNeed</td>
<td>dial the number, find the number, look up things online, to pick up the phone, ...</td>
<td>to have a sick animal, to get doctor's phone number</td>
</tr>
<tr>
<td>PersonX comes back, PersonX comes to PersonY’s house</td>
<td>PersonX comes to PersonY’s house</td>
<td>oWant</td>
<td>to greet PersonX, to hug him, to help him relax, to eat out, to invite PersonX inside, ...</td>
<td>to go eat, to have dinner, to talk to PersonX</td>
</tr>
<tr>
<td></td>
<td></td>
<td>xIntent</td>
<td>see their family, to get home, to sleep, see their family, to attend some competition, ...</td>
<td>have a break from learning, to attend the wedding</td>
</tr>
<tr>
<td></td>
<td></td>
<td>xReact</td>
<td>cozy, happy, nostalgic, relaxed</td>
<td>drink, ready to eat, sleepy</td>
</tr>
</tbody>
</table>
Concept-aided Situational Commonsense Modeling

• A more abstract view may help the model to learn?
  • Augment ATOMIC with abstract knowledge
    • Especially with limited data
  • Use the ATOMIC subset that constitute the base of events in annotated triples
    • Mix with annotated or the corresponding automatically-built triples
    • Further finetune on ATOMIC

<table>
<thead>
<tr>
<th>BLEU-2</th>
<th>GPT2-base</th>
<th>GPT2-medium</th>
</tr>
</thead>
<tbody>
<tr>
<td>Baseline (COMET)</td>
<td>17.7</td>
<td>19.6</td>
</tr>
<tr>
<td>+Conceptualization (Human)</td>
<td>20.6</td>
<td>23.5</td>
</tr>
<tr>
<td>+Conceptualization (Auto)</td>
<td>19.3</td>
<td>21.0</td>
</tr>
<tr>
<td>+Conceptualization (Both)</td>
<td>19.0</td>
<td>22.9</td>
</tr>
</tbody>
</table>
Conclusion and Future Work

• A framework for machine conceptualization is formulated and implemented
  • A dataset for validity of conceptualization is annotated
  • Heuristic rules and neural models to generate and verify conceptualization are developed
  • A large scale abstract CKG is inferred
    • 70K abstract events and 2.9M abstract triples

• Future work
  • Better models
  • More downstream tasks
  • Integrating more data, e.g., ATOMIC-10X

https://github.com/HKUST-KnowComp/atomic-conceptualization