Commonsense for Machine Intelligence: Text to Knowledge and Knowledge to Text

Gerard de Melo, Niket Tandon and Aparna S. Varde

Rutgers University
Allen Institute for Artificial Intelligence
Montclair State University

Research Tutorial at ACM CIKM 2017
Singapore, Nov 6 to 10, 2017
Tutorial Presenters

Gerard de Melo, Assistant Professor, Rutgers University, New Brunswick, NJ, USA

Niket Tandon, Research Scientist, Allen Institute for Artificial Intelligence, Seattle, WA, USA

Aparna S. Varde, Associate Professor, Montclair State University, Montclair, NJ, USA
Tutorial Agenda

Part 1: Acquiring Commonsense Knowledge

Part 2: Detecting and Correcting Odd Collocations in Text

Part 3: Applications and Open Issues
Part 1: Acquiring Commonsense Knowledge

Commonsense for Machine Intelligence: Text to Knowledge and Knowledge to Text
Part 1: Acquiring Commonsense Knowledge

• introduction
  • introduction to csk
  • csk unimodal and multimodal kbs

• csk representation
  • discrete and continuous representations
  • multimodal continuous representations

• acquisition methods
  • different levels of supervision and modalities
  • from facts to rules

• csk evaluation
  • explicit evaluation techniques: sampling, turked
  • challenge sets and problems in text and vision
What is commonsense?
Questions from Aristo challenge (allennai.org/data)

Common knowledge about things in the world, their associations, and interactions. Commonsense knowledge is mostly location and culture independent.
Machines cannot reason like humans, because they lack commonsense.

For roller-skate race, what is the best surface? (A) sand (B) grass (C) blacktop

Picture depicts food web. Arrow indicates consumes. If frogs die, raccoons won’t get food and die-- so their population will decrease.

Humans
Knowledge Gap

Roller skate is best on a smooth surface. Blacktop surfaces are shiny and shiny surfaces are smooth.

grass – related to – field – race, so “grass”
blacktop is not related to race.
What about the Knowledge graphs?

Machines can surpass most humans on Encyclopedic knowledge about popular “named entities”

Machines cannot surpass any human on commonsense knowledge about “common nouns”
I touched a table that was hard.

roots absorb water $\rightarrow$ water is at the roots

Less, implicitly expressed
[Tandon et. al Ngram workshop 2010]

Reporting Bias
[Parikh EACL 2017, Gordon et. al AKBC 2013]

Multimodal
[Tandon et. al AAAI 2016]
Elusive

Less, implicitly expressed
Reporting Bias
Multimodal

Contextual

Depends on the context
[Liu et. al 2004]
<table>
<thead>
<tr>
<th>Implicitness Increases</th>
<th>Properties</th>
<th>Relationships</th>
<th>Interactions</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Properties</td>
<td>Taxonomic</td>
<td>Actions</td>
</tr>
<tr>
<td></td>
<td>Theories</td>
<td>Spatial</td>
<td>Procedural</td>
</tr>
<tr>
<td></td>
<td>Emotions</td>
<td>Structures</td>
<td>Behaviors</td>
</tr>
</tbody>
</table>

Commonsense knowledge types in multiple modalities

- Implicitness increases: The implicitness of knowledge increases from top to bottom.
- Visual modality dominates: The visual modality dominates from left to right.
Commonsense Knowledge types

- **Properties**
  - Shiny surfaces are hard

- **Taxonomic Relationships**
  - Climbers are humans

- **Actions**
  - Climbing involves ropes

- **Theories**
  - Friction increases → speed decreases

- **Spatial Relationships**
  - Rock on mountain, by the river

- **Procedural Interactions**
  - How to fix a tire

- **Emotions**
  - Climbing evokes adventure

- **Structures**
  - Wheel is part of a bike

- **Behaviors**
  - During storms, sea becomes violent

Motivated by Peter Clark’s talk at AI2
### Commonsense Knowledge KBs

<table>
<thead>
<tr>
<th>Implicitness Increases</th>
<th>Properties</th>
<th>Relationships</th>
<th>Interactions</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td><strong>Properties</strong>&lt;br&gt;Cyc, ConceptNet, WebChild</td>
<td><strong>Taxonomic</strong>&lt;br&gt;WordNet, UWN, Cyc, LEVAN</td>
<td><strong>Actions</strong>&lt;br&gt;Cyc, ConceptNet, WebChild, VisualGenome, LEVAN</td>
</tr>
<tr>
<td></td>
<td><strong>Theories</strong>&lt;br&gt;Cyc</td>
<td><strong>Spatial</strong>&lt;br&gt;Cyc, ConceptNet, WebChild, NEIL, RoboBrain, VisualGenome</td>
<td><strong>Procedural</strong>&lt;br&gt;WebChild/HowToKB</td>
</tr>
<tr>
<td></td>
<td><strong>Emotions</strong>&lt;br&gt;SenticNet, ConceptNet, WebChild</td>
<td><strong>Structures</strong>&lt;br&gt;Cyc, ConceptNet, WebChild, NEIL</td>
<td><strong>Behaviors</strong>&lt;br&gt;Cyc</td>
</tr>
</tbody>
</table>

- Visual modality dominates

---

Implicitness increases along the vertical axis, and the modalities dominate along the horizontal axis.
Note: There are many more taxonomies such as Microsoft ProBase.
Cyc

[Lenat 1995]
ConceptNet

[Liu et al. 2004]
Aristo tuple KB

<table>
<thead>
<tr>
<th>Tree</th>
<th>trunk</th>
<th>bark</th>
<th>vacuole</th>
<th>cell membrane</th>
<th>cell</th>
<th>nucleus</th>
<th>treetop</th>
</tr>
</thead>
<tbody>
<tr>
<td>has-part</td>
<td>xylem</td>
<td>stump</td>
<td>cytoplasm</td>
<td>tree branch</td>
<td>corpus</td>
<td>plasma membrane</td>
<td></td>
</tr>
<tr>
<td></td>
<td>section</td>
<td>leaf node</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>feature</td>
<td>trunk</td>
<td>massive trunk</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>carry</td>
<td>leaf</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>occupy</td>
<td>habitat</td>
<td>rocky habitat</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>use</td>
<td>photosynthesis</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Reduced expressivity

Syntactic structures

OpenIE

ConceptNet

PropBank

Semantic expressivity

Generic

HowToKB

Fillmore’76

Schank’75

Minsky’74

VerbNet

FrameNet

Domain specific

Knowlywood
**mountain**: a land mass that projects well above its surroundings; higher than a hill

<table>
<thead>
<tr>
<th>TYPE OF</th>
<th>natural elevation</th>
</tr>
</thead>
<tbody>
<tr>
<td>PHYSICAL PROPERTIES</td>
<td>large, high, heavy, cold, hard More</td>
</tr>
<tr>
<td>ABSTRACT PROPERTIES</td>
<td>elegant, old, safe, holy, risky More</td>
</tr>
<tr>
<td>COMPARABLES</td>
<td>mountain, hill, mountain, mount, mountain, high hill, valley, mountain More</td>
</tr>
<tr>
<td>HAS PHYSICAL PARTS</td>
<td>mountain peak, mountainside, slope, tableland, hill More</td>
</tr>
<tr>
<td>HAS SUBSTANCE</td>
<td>mixture, metallic element, material, page, wood More</td>
</tr>
<tr>
<td>IN SPATIAL PROXIMITY</td>
<td>coast, tunnel, lake, sea, river More</td>
</tr>
<tr>
<td>ACTIVITIES</td>
<td>climb mountain, cross mountain, move mountain, see mountain, ascend mountain</td>
</tr>
<tr>
<td>Regions</td>
<td>Attributes</td>
</tr>
<tr>
<td>---------</td>
<td>------------</td>
</tr>
<tr>
<td>A wall on the side of a building</td>
<td>leaf is green and yellow</td>
</tr>
<tr>
<td>green leaf with yellow spots</td>
<td>stems is green</td>
</tr>
<tr>
<td>yellowish green stems</td>
<td>light is bright</td>
</tr>
<tr>
<td>bright light reflecting on leaves</td>
<td>leaf is dark</td>
</tr>
<tr>
<td>a dark green leaf with brown edges</td>
<td>leaf is green</td>
</tr>
<tr>
<td>a large green leaf</td>
<td>leaf is brown edge</td>
</tr>
<tr>
<td>a large dark green leaf</td>
<td>leaf is large</td>
</tr>
<tr>
<td></td>
<td>leaf is dark green</td>
</tr>
<tr>
<td></td>
<td>leaf is yellow, brown, and green</td>
</tr>
<tr>
<td></td>
<td>pale flower is pale</td>
</tr>
</tbody>
</table>
object (horse), scene (kitchen), event (xmas), action (walking)
Part 1: Acquiring Commonsense Knowledge

- introduction
  - introduction to csk
  - csk unimodal and multimodal kbs
- **csk representation**
  - discrete and continuous representations
  - multimodal continuous representations
- acquisition methods
  - different levels of supervision and modalities
  - from facts to rules
- **csk evaluation**
  - explicit evaluation techniques: sampling, turked
  - challenge sets and problems in text and vision
Classical deductive reasoning requires formal representations for syllogisms. Cyc’s microtheories is a classic example of formal representations.

All trees are plants
(#$genls #$Tree-ThePlant #$Plant) ;

For roller-skate race, what is the best surface?
(A) sand (B) grass (C) blacktop

blacktop surfaces are shiny
shiny surfaces are smooth
smooth surfaces have less friction
less friction speeds up roller-skates
you win race when fastest

Problems
• Require perfect knowledge for heavy duty deductive inference, and,
• Far from natural language, query must be translated to this representation.
ConceptNet has a string representation, over a fixed set of relations.

+ Admits an open vocabulary

- difficult knowledge retrieval and generalization, especially under reporting bias.
  - plants, absorb, solar energy
  - trees, take in, sunlight
Aristo TupleKB and WebChild use ILPs to cluster for structure by maximizing for coherence, overcoming reporting bias by aggregating frequencies.

\[ \text{<wet wood, softer than, dry wood>} \equiv \text{<dry wood, harder than, wet wood>} \]

\[
\begin{align*}
\text{max (triple internal coherence + sense frequency)} & \equiv \text{in WN space} & \equiv \text{max (triple internal coherence + sense frequency)} \\
\text{<wet wood-n-1, softer-a-1 than, dry wood-n-1>} & \equiv \text{<dry wood-n-1, harder-a-1 than, wet wood-n-1>} &
\end{align*}
\]
Higher arity tuples, such as OpenIE [Etzioni et. al 2011]: 
<s, p, o, location, time>

further structured into frame based representations to retain more context.

+ useful when a relation admits multiple values
+ allows for maintaining top-k values by salience
Continuous representations.

Why continuous representation?

Knowledge retrieval over discrete representations suffers from linguistic variations.

Using matrix and tensor factorizations [Bordes et. al 2014], continuous representations in the embedding space has made progress for Encylopedic KB completion. [Jain et. al 2017] found that progress across popular data sets does not generalize.

A recent approach injects knowledge in a memory cell, KB-LSTMs [Yang et. al 2017] such that the hidden vector learned per word is biased with KB context. Gains were limited, but this is a promising direction.
If \((e_i, r_k, e_j)\) holds, then
\((e_i, r_k, e_j') \neq e_j\) is -ve

A. Honnold, bornIn, US
A. Honnold, bornIn, UK

CKB: assumption fails

Encyclopedic Knowledge

.commonsense

EKBs have several functional relations hence the assumption holds.

- Functional
- Non-functional

Classes generalize properties of instances
DeViSE [Frome et. al 2013] - continuous multimodal representation

Trains a joint embedding model of both images and labels (text), with non-linear mappings from image features to the embedding space.

The joint model is initialized with parameters pre-trained at the lower layers of traditional image, text models.
Part 1: Acquiring Commonsense Knowledge

• introduction
  • introduction to csk
  • csk unimodal and multimodal kbs

• csk representation
  • discrete and continuous representations
  • multimodal continuous representations

• acquisition methods
  • different levels of supervision and modalities
  • from facts to rules

• csk evaluation
  • explicit evaluation techniques: sampling, turked
  • challenge sets and problems in text and vision
Sources of acquisition

- Authored
  - experts, crowd, games
- Text
  - Web, ngrams, books, tables, moviescripts, Wikipedias
- Images:
  - Flickr annotations, clip art, real images, videos
Acquisition methods

• different levels of supervision and modalities
• from facts to rules
• WordNet is a lexical DB with taxonomies of nouns, verbs.
• Linguistics authored WN DAG, to a very fine-grained level.
• Sense orderings in WN can be uncommon with freq. statistics from an annotated corpora.
  • the first sense of “tiger” is ... a fierce man.
• Cyc is being built for 30 years by several experts, still about a million facts

• Cyc is a collection of microtheories partitioned by domain

• CycL is a language in which these microtheories can be operated, to perform very powerful deductive reasoning.

• The reasons for the failure of Cyc has been that:
  • it is far from language: Input query must to be translated into CycL first.
  • it is very difficult to know which microtheory is applicable for a scenario.
  • deductive reasoning is powerful, but brittle: commonsense tends to be contextual
• Step 1: Collect OMCS corpus of CSK sentences
  • Templatized—closed set of relations.
  • ___ can be used to ___ expected to filled as e.g., pen, write.

• Step 2: Relation extraction becomes a regex match
  • Post-processed strings to minimize noise, scores = agreements.
  • Negations are a highlight of ConceptNet.

• Step 3: Repeated Step 1, 2 for multiple languages
  • These are independent efforts.

• Step 4: Vectorized the factorized concepts in concept x attributes matrix
  • Recently, ConceptNet is retrofitted to Glove embeddings.

• Note: ConceptNet5 onwards is a mix of non-commonsense and commonsense knowledge. E.g., >99% of the ~5 million part-whole relations are geo-locations.
  • Marina Bay is part of Asia – whereas, one would expect leaf is part of a plant.
• Player 1 (describer) tells a secret word to player 2 (guesser) by filling in provided templates.
  • “It is a type of ___”,
  • “About the same size as ___”.
• Correct guesses lead to a fact.
• Note: Engagement in commonsense games is difficult.
Pattern based methods – distant supervision.

• For a fixed set of relations, works in two steps:
  • Step 1: Find patterns that cover seeds facts.
  • Step 2: Apply patterns to find facts.

• Two considerations:
  • Typed (e.g., POS tagged) patterns for better accuracy
  • Pattern scoring is very important when relations are related (as in CSK)

• Usually:
  • Semantic drift after initial round(s).
  • Curating after each round leads to higher accuracy.
Pattern based methods – distant supervision [Tandon et. al 2016].

Noisy patterns must be pruned

* are essential parts of *
* on sale along with *

PE are essential parts of PE
seat, cycle
dollars, life
snake, machine

Type checking always helps

Physical entity, abstract entity help tell apart easily confused relations

\[
\sigma(a_k) = \frac{e^{supp(a_k)}}{1 + e^{supp(a_k)}} \quad e^{str(a_k)}
\]

Score of a candidate \(a_k\) (either pattern or assertion)

Seed Support: many distinct seed matches implies high score

Strength: Penalize semantic drift of candidates matching seeds of diff. relations

<table>
<thead>
<tr>
<th>Domain (r)</th>
<th>Relation: r</th>
<th>Range (r)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Physical Entity</td>
<td>Physical part of (P) wheel, cycle</td>
<td></td>
</tr>
<tr>
<td>Physical + Abstract Entity</td>
<td>Member of (M) cyclist, team</td>
<td></td>
</tr>
<tr>
<td>Substance Entity</td>
<td>Substance of (S) rubber, wheel</td>
<td>Physical Entity</td>
</tr>
</tbody>
</table>
• When the amount of training data is limited/ expensive
• The space is naturally clusterable
• Well studied machine learning methods to leverage.
**Task:** Input = web corpus, bootstrapping patterns to extract adjective noun pairs

e.g., <summit, crisp>

**Output:** triples $<w^1_s, r, w^2_s>$

**disambiguated n**

1.)

2.)

3.) ...

**disambiguated a**

1.)

2.)

3.) ...

**fine-grained relations:** $r \in R$

- hasAppearance
- hasSound
- hasTaste
- hasTemperature
- evokesEmotion

- hasSound
- evokesEmotion

- hasSound
- evokesEmotion

**Semi-supervised learning in WebChild [Tandon et. al 2014]**

Task: Input = web corpus, bootstrapping patterns to extract adjective noun pairs
e.g., <summit, crisp>
Semi-supervised learning in WebChild [Tandon et. al 2014]

Task: Input = web corpus, bootstrapping patterns to extract adjective noun pairs
e.g., <summit, crisp>

Output: triples

\[
\langle w_1^s, r, w_2^a \rangle
\]

**disambiguated n**

1.)

2.)

3.)

... 

**disambiguated a**

1.)

2.)

3.)

... 

**fine-grained relations:**

\[ r \in R \]

- hasAppearance
- hasSound
- hasTaste
- hasTemperature
- evokesEmotion

**Assertion inference**

**Domain inference**

**Range inference**
An instance of the problem: $\text{range}(r = \text{hasTemperature})$
An instance of the problem: $\text{range}(r)$

\[
\tau_{AA}[a_1, a_2] = \alpha OTO + (1 - \alpha)PTP
\]

\[
\tau_{AA}[a^i, a^j] = \beta \text{hirst}[a^i, a^j] + (1 - \beta)G^T G
\]

\[
\phi[a, a^i] = \gamma \frac{1}{1+1} + (1 - \gamma)O^T G
\]

<table>
<thead>
<tr>
<th></th>
<th>summit</th>
<th>mountain</th>
<th>dancer</th>
</tr>
</thead>
<tbody>
<tr>
<td>cold</td>
<td>20</td>
<td>50</td>
<td>3</td>
</tr>
<tr>
<td>hot</td>
<td>30</td>
<td>40</td>
<td>10</td>
</tr>
<tr>
<td>crisp</td>
<td>15</td>
<td>15</td>
<td>1</td>
</tr>
</tbody>
</table>

\[
O_{i,j} : \ freq(\text{noun}_i, \text{adj}_j)
\]

\[
P_{i,j} : \ #\text{patt}(\text{noun}_i, \text{adj}_j)
\]

\[
G_{i,j} : \ freq(\text{adj}_i, \text{glossword}_j)
\]
Label propagation for graph inference, given few seeds.
- Label per node = in/not in range of hasTemperature

Similar nodes
Similar labels
But, limited training data
Label Propagation: Loss function [Talukdar et. al 2009]

\[
\mu_1 \sum_v p_{\nu}^{inj} \sum_l (Y_{vl} - \hat{Y}_{vl})^2 + \mu_2 \sum_{v,u} p_{\nu}^{cont} W_{vu} \sum_l (Y_{vl} - \hat{Y}_{ul})^2 + \mu_3 \sum_{vl} (\hat{Y}_{vl} - R_{vl})^2
\]
<table>
<thead>
<tr>
<th>Relation</th>
<th>Range</th>
<th>Domain</th>
<th>Assertions</th>
</tr>
</thead>
<tbody>
<tr>
<td>hasTaste</td>
<td>sweet\textsubscript{1_a}</td>
<td>strawberry\textsubscript{1_n}</td>
<td>(chocolate\textsubscript{1_n}, creamy\textsubscript{2_a})</td>
</tr>
<tr>
<td></td>
<td>hot\textsubscript{9_a}</td>
<td>chili\textsubscript{1_n}</td>
<td>(pizza\textsubscript{1_n}, delectable\textsubscript{1_a})</td>
</tr>
<tr>
<td></td>
<td>sour\textsubscript{2_a}</td>
<td>salsa\textsubscript{1_n}</td>
<td>(salsa\textsubscript{1_n}, spicy\textsubscript{2_a})</td>
</tr>
<tr>
<td></td>
<td>salty\textsubscript{3_a}</td>
<td>sushi\textsubscript{1_n}</td>
<td>(burger\textsubscript{1_n}, tasty\textsubscript{1_a})</td>
</tr>
<tr>
<td></td>
<td>lemony\textsubscript{1_a}</td>
<td>java\textsubscript{2_n}</td>
<td>(biscuit\textsubscript{2_n}, sweet\textsubscript{1_a})</td>
</tr>
<tr>
<td></td>
<td>triangular\textsubscript{1_a}</td>
<td>leaf\textsubscript{1_n}</td>
<td>(palace\textsubscript{1_n}, domed\textsubscript{1_a})</td>
</tr>
<tr>
<td></td>
<td>meandering\textsubscript{1_a}</td>
<td>circle\textsubscript{1_n}</td>
<td>(table\textsubscript{2_n}, flat\textsubscript{1_a})</td>
</tr>
<tr>
<td></td>
<td>crescent\textsubscript{1_a}</td>
<td>ring\textsubscript{8_a}</td>
<td>(jeans\textsubscript{2_n}, tapered\textsubscript{1_a})</td>
</tr>
<tr>
<td></td>
<td>obtuse\textsubscript{2_a}</td>
<td>egg\textsubscript{1_n}</td>
<td>(tv\textsubscript{2_n}, flat\textsubscript{1_a})</td>
</tr>
<tr>
<td></td>
<td>tapered\textsubscript{1_a}</td>
<td>face\textsubscript{1_n}</td>
<td>(lens\textsubscript{1_n}, spherical\textsubscript{2_a})</td>
</tr>
</tbody>
</table>
Semi-supervised learning for knowledge frame

{Climb up a mountain, Hike up a hill}

- **Participants**: climber, boy, rope
- **Location**: camp, forest, sea, shore
- **Time**: daylight, holiday
- **Visuals**: Go up an elevation

**Previous activity**: Get to village

**Parent activity**: Climbing

**Next activity**: Drink water
Judicious choice of multimodal dataset

May contain events or activities but varying granularity and no visuals. No clear scene boundaries.

Hollywood narratives are easily available and meet the desiderata.

EXT. SMALL MOUNTAIN--DAY

Wichita charges up the rockage of a small mountain-hill-type thing. The image repeats itself over and over--each time Wichita is more sweaty, gasping, sneering.

Wichita (V.O.)
The rules forbid anyone from the climbing the camp's mountain.

align via subtitles with approximate dialogue similarity
State of the art WSD customized for phrases

Syntactic and semantic role semantics from VerbNet

Output Frame

Agent: man.1
Action: shoot.4
Patient: video.1

Thing/inanimate via WN
Semantic parsing of scripts

Graph inference

- Climb up a mountain
  - Participants: climber, rope
  - Location: summit, forest
  - Time: day

- Hike up a hill
  - Participants: climber
  - Location: sea shore
  - Time: holiday

- Go up an elevation

- Reach top

Similar

Parent

Temporal
Semantic parsing of scripts → Graph construction → Knowlywood Activity Knowledge Base
Visual Genome – crowdsourced visual KB

- Crowdsourced on mechanical turk.
- Contains bounding boxes annotated with relationships.
Learn commonsense through visual abstraction [Vedantam et al. 2015]

- "look at" and "want" not semantically sim.
- Input = {squirrel (s), looks at (p), nuts (o)}
- Output = plausible/ not

\[
\beta \cdot \text{text plausibility w.r.t. train} + (1-\beta) \cdot \text{visual plausibility w.r.t. train}
\]

\[
\beta \cdot \text{vector space sim. } V(s_1, s_1') + V(p_1, p_1') + V(o_1, o_1') + (1-\beta) \cdot \text{Object features + Scene features + Interaction features}
\]

Unsupervised text visuals
Manual

Train data collection

Demonstrate this relation:

Name the objects that you selected to participate in this relation (as brief as possible):
Learning commonsense directly through images [Shrivastava et al. 2014]

- **Sparrow** is a kind of/looks similar to **bird**
- **Eye** is a part of **Baby**

**Object - Scene**

- **Helicopter** is found in **Airfield**
- **Ferris wheel** is found in **Amusement park**
(0) Seed Images

(1) Subcategory Discovery

(2) Train Models

(3) Relationship Discovery

(4) Add New Instances

Learned relationships:
- Keyboard is a part of Desktop Computer
- Monitor is a part of Desktop Computer
- Television looks similar to Monitor
AMIE is a system that mines Horn rules from a KB.
- \( \text{hasChild}(z, y) \land \text{married}(x, z) \Rightarrow \text{hasChild}(x, y) \)
- Rules can be used to make predictions and in reasoning.
- Performs exhaustive top-down search based on partial closed world assumption, minimum support threshold.

Scoring of candidate rules based on PCA confidence: the ratio of predicted positive examples, out of all predicted examples.
Part 1: Acquiring Commonsense Knowledge

• introduction
  • introduction to csk
  • csk unimodal and multimodal kbs

• csk representation
  • discrete and continuous representations
  • multimodal continuous representations

• acquisition methods
  • different levels of supervision and modalities
  • from facts to rules

• csk evaluation
  • explicit evaluation techniques: sampling, turked
  • challenge sets and problems in text and vision
Evaluation of acquired knowledge

• Commonsense relations are typically non-functional (e.g., location of a car – roads, parking lot, highway)

• Most approaches (e.g., Tandon 2014) are based on a manual assessment including from turkers, where the task is to identify if a relation “can hold”.

• The non-functional slice of encyclopedic knowledge (e.g., childrenOf) does not require “prominent” values. Commonsense acquisition must capture the most salient answers, however. Newer intrinsic evaluations have started asking “What is the 2-3 most prominent values for this concept, under a certain relation).

• Estimating recall remains an open problem because we do not know the possible commonsense knowledge. Dalvi et. al 2017 propose completeness w.r.t. a corpus, of prominent facts from a book.
Evaluation of acquired knowledge

• Verification of commonsense knowledge from text, by detecting inconsistencies based on manually specified rules such as transitivity:
  • e.g., spoke part of wheel \(\rightarrow\) wheel part of cycle \(\rightarrow\) spoke part of cycle

• Due to reporting bias, visual verification is becoming popular. Visual verification is not an option for Encyclopedic knowledge.

• Tandon et. al 2016 described how part-whole relationships (such as, seat is part of cycle) can be verified using images or more scalably using image tags.
  • **cycle**, fun, trip, go, **seat**, niket
    seat is visible part of cycle
Visual verification using low level image features [Sadeghi et. al 2015]

Input: fish (bear, salmon)

Images Retrieved
- bear
- salmon
- bear fishing
- fishing salmon
- bear fishing salmon

Trained Detection Models
- bear (S)
- salmon (O)
- bear fishing (SV)
- fishing salmon (VO)
- bear fishing salmon (SVO)

Output: bear fishing salmon

MPE(R)
Evaluation of acquired knowledge

Commonsense for better concept-concept similarity, but this evaluation is limited by scale and subjectivity (Tandon et al. 2014)

<table>
<thead>
<tr>
<th>Top 10 adjectives</th>
<th>universal, magnetic, small, ornamental, decorative, solid, heavy, white, light, cosmetic</th>
</tr>
</thead>
<tbody>
<tr>
<td>Top 5 expansions</td>
<td>wall mount, mounting bracket, wooden frame, carry case, pouch</td>
</tr>
</tbody>
</table>
A number of disparate large-scale annotated challenge sets for commonsense exist.

However, unlike some standard tasks like reading comprehension or object detection, these have not been the heralds of commonsense knowledge as they require a reasoner which makes the evaluation subjective.

More end to end or task-oriented evaluation of commonsense knowledge is needed.
The StoryCloze dataset [Mostafazadeh et. al 2016] requires predicting the conclusion of a story. However, elimination of answer choices is an effective approach for this dataset, rendering this task inference easy.

<table>
<thead>
<tr>
<th>Context</th>
<th>Right Ending</th>
<th>Wrong Ending</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sammy’s coffee grinder was broken. He needed something to crush up his coffee beans. He put his coffee beans in a plastic bag. He tried crushing them with a hammer.</td>
<td>It worked for Sammy.</td>
<td>Sammy was not that much into coffee.</td>
</tr>
<tr>
<td>Gina misplaced her phone at her grandparents. It wasn’t anywhere in the living room. She realized she was in the car before. She grabbed her dad’s keys and ran outside.</td>
<td>She found her phone in the car.</td>
<td>She didn’t want her phone anymore.</td>
</tr>
</tbody>
</table>
Challenge sets and problems for commonsense

**MovieQA** [Tapaswi et. al 2016] poses QA over movie video clips, plots, subtitles, scripts, and DVS [Rohrbach et. al 2015].

Inference:
- Easy: Text
- Hard: Visual

**Quiz**
- **What is the Matrix?**
  - A: A shared simulation of the world
  - A: A group of robots
  - A: A human body
  - A: A set of numbers stored as a table

- **Who kills Neo in the Matrix?**
  - A: Smith kills Neo
  - A: Trinity kills Neo
  - A: Morpheus kills Neo after he realizes that Neo is not the one

- **Why does Cypher betray Morpheus?**
  - A: In exchange for a comfortable life
  - A: In exchange for money
  - A: Because he is threatened by Agent Smith

- **How does the movie end?**
  - A: With Neo flying into the sky
  - A: With the Machines chasing after Neo
  - A: We see Mr. Smith torture Morpheus
Mariano fell with a crash and lay stunned on the ground. Castello instantly kneeled by his side and raised his head. His head: Mariano/ Castello?

### Stories (Winograd [Liu et. al 2016], SemEval 2018 Task 11[Ostermann 2017])

Consider the following reading text from the planting a tree scenario...

My backyard was looking a little empty, so I decided I would plant something. I went out and bought tree seeds. I found a spot in my yard that looked like it would get enough sunshine. There, I dug a hole for the seeds. Once that was done, I took my watering can and watered the seeds.

... and the following questions on the text.

A. Why was the tree planted in that spot?
   1. to get enough sunshine
   2. there was no other space

B. What was used to dig the hole?
   1. a shovel
   2. their bare hands

C. Who took the watering can?
   1. the grandmother
   2. the gardener
For roller-skate race, what is the best surface?
(A) sand (B) grass (C) blacktop
Challenge sets and problems for commonsense

QA (TQA [Kembhavi et. al 2016], Aristo [Clark 2014], FVQA)

Aristo challenge  allenai.org/data

If all the frogs died, the raccoon population would most likely
(A) decrease (B) increase (C) remain the same
Challenge sets and problems for commonsense

TQA/ Aristo, FVQA [Wang et. al 2016] includes supporting facts from KB

Q: What things in this image are eatable?
A: Apples

Q: What is the order of the animal described in this image?
A: Odd toed ungulate

Q: What thing in this image is helpful for a romantic dinner?
A: Wine
Uses an LSTM and a data-driven approach to learn the mapping of images/questions to queries. Uses DBpedia, ConceptNet and WebChild.

### KB | Predicate | #Facts | Examples
---|---|---|---
DBpedia | Category | 35152 | (Wii, Category, VideoGameConsole)
| RelatedTo | 79789 | (Horse, RelatedTo, Zebra), (Wine, RelatedTo, Goblet)
| AtLocation | 13683 | (Bikini, AtLocation, Beach), (Tap, AtLocation, Bathroom)
| IsA | 6011 | (Broccoli, IsA, GreenVegetable)
| CapableOf | 5837 | (Monitor, CapableOf, DisplayImages)
| UsedFor | 5363 | (Lighthouse, UsedFor, SignalingDanger)
| Desires | 3358 | (Dog, Desires, PlayFrisbee), (Bee, Desires, Flower)
| HasProperty | 2813 | (Wedding, HasProperty, Romantic)
| HasA | 1665 | (Giraffe, HasA, LongTongue), (Cat, HasA, Claw)
| PartOf | 762 | (RAM, PartOf, Computer), (Tail, PartOf, Zebra)
| CreatedBy | 96 | (Bread, CreatedBy, Flour), (Cheese, CreatedBy, Milk)

ConceptNet

WebChild

| Small, Better, | Slower, Bigger, | Taller, ... | 38576 | (Motorcycle, Smaller, Car), (Apple, Better, VitaminPill), (Train, Slower, Plane), (Watermelon, Bigger, Orange), (Giraffe, Taller, Rhino) |
Summary of Part 1

**CSK types**

- Properties: Con, ConceptNet, WebChild
- Taxonomic: Weisi, LAVI, Con, LEVAN
- Actions: Con, ConceptNet, WebChild, Visual Genome, LEVAN
- Theories: Con
- Spatial: Con, ConceptNet, WebChild, RES, Robotbrain
- Procedural: WebChild
- Emotions: SemCor, ConceptNet, WebChild
- Structures: Con, ConceptNet, WebChild, NEIL
- Behaviors: Con

**CSK representation**

- Structured
- Continuous
- Discrete
- Natural lang

**CSK acquisition**

- Text
- Visuals
- Unsupervised
- Manual

**CSK evaluation**

- Manual evaluation
- Intrinsic
- Automated evaluation
- Extrinsic
Future directions

1. Modeling implicit states through simulations:
   - Physical CSK – implicit states: roots absorb water \(\rightarrow\) water is at roots
   - Social CSK– man: I broke up with Jenny, robot: do you want to go out with her?

2. Multimodal CSK (vision, text, and more e.g. audio for continuous learning)
   - Can commonsense be derived only from videos, how do we know what relations to focus?

3. Salient and concise KBs: efficient computations, quality control.
   - dogs have eyes, cats have eyes, squirrels have eyes.. vs. mammals have eyes.

4. Better evaluation metrics:
   - KB comprehensiveness as a metric– task independent
   - Extrinsic evaluations to continuously track progress
References

• Chambers, N., & Jurafsky, D. (2010). A Database of Narrative Schemas. LREC.
• Chang, A. X., Savva, M., Manning, C. (2014). Learning Spatial Knowledge for Text to 3D Scene Generation. EMNLP.
• Chen, J., Tandon, N., & De Melo, G. (2016). Neural word representations from large-scale commonsense knowledge. WI.
• Clark, P. (2014) Elementary School Science and Math Tests as a Driver for AI: Take the Aristo Challenge! IAAI.
• de Melo, G., & Weikum, G. (2009). Towards a universal wordnet by learning from combined evidence. CIKM.
• Li, X., Vilnis, L., & McCallum (2017). Improved Representation Learning for Predicting Commonsense Ontologies


• Pichotta, K., & Mooney, R. J. (2016). Learning Statistical Scripts With LSTM Recurrent Neural Networks (AAAI)


• Tapaswi, M., Bäuml, M., & Stiefelhagen, R. (n.d.). Book2Movie: Aligning Video scenes with Book chapters.