Imperfect Knowledge

Dave Raggett dsr@w3.org

DKG-22, 16 February 2022

This work has been supported through funding for the TERMINET project from the European Union's Horizon 2020 research and innovation programme under grant agreement No 957406.

Reasoning with Knowledge Graphs

- Whilst knowledge graphs claim to capture knowledge there is very little attention currently to automated reasoning
  - One exception is inheritance down class hierarchies
- Application logic is instead embedded in application code
  - This makes it hard to understand and to update
- Knowledge presumes reasoning and is otherwise just information!
- It is high time to focus on automated reasoning for human-machine cooperative work

Everyday Knowledge is Imperfect

- In the real world, knowledge is distributed and imperfect
- We are learning all the time, and revising our beliefs and understanding as we interact with others
- Imperfect in the sense of **uncertain**, **incomplete** and **inconsistent**
- Conventional logic fails to cope with this challenge
- The same is true for statistical approaches, e.g. Bayesian inference, due to the lack of the required statistics
- Evolution has equipped humans with the means to deal with this
- However, not everyone is rational, and some lack sound judgement*

* Something exploited by unscrupulous politicians and advertisers

Plausible Reasoning and Argumentation

- People have studied the principles for plausible arguments since the days of classical Greece, e.g. Carneades and his guidelines for argumentation
- There has been a long line of philosophers working on this since then, including Locke, Bentham, Wigmore, Keynes, Wittgenstein, Pollock and many others
- Plausible reasoning is *everyday reasoning*, and the basis for legal, ethical and business discussions
- Researchers in the 20th century were sidetracked by the seductive purity of mathematical logic, and more recently, the magic of deep learning
- It is now time to exploit plausible reasoning with imperfect knowledge for human-machine collaboration using distributed knowledge graphs
- Enabling computers to analyse, explain, justify, expand upon, and argue

Plausible Inferences

- Consider $A \Rightarrow B$, which means if A is true then B is true. If A is false then B may be true or false. If B is true, we still can't be sure that A is true, but if B is false then A must be false
- A more concrete example: *if it is raining then it is cloudy*
  - This can be used in both directions: Rain is more likely if it is cloudy, likewise, if it is not raining, then it might be sunny, so it is less likely that it is cloudy*
- In essence, plausible reasoning draws upon prior knowledge as well as on the role of analogies and consideration of examples, including *precedents*
- Mathematical proof is replaced by reasonable arguments, both for and against a premise, along with how these are assessed
  - In court cases: arguments are laid out by the Prosecution and Defence, and assessed by the Judge and Jury

* Note use of qualitative terms in lieu of quantitive statistics

Plausible Reasoning

During the 80’s Alan Collins and co-workers developed a theory of plausible reasoning* based upon recordings of how people reasoned. They found that:

- There are several categories of inference rules that people commonly use to answer questions
- People weigh the evidence that bears on a question, both for and against
- People are more or less certain depending on the certainty of the premises, the certainty of the inferences and whether different inferences lead to the same or opposite conclusions
- Facing a question for which there is an absence of directly applicable knowledge, people search for other knowledge that could help given potential inferences

* See: Collins & Michalski (1988) and subsequent extensions by Burstein, Collins and Baker (1991)

Proof of Concept

- A web-based demo that provides a simple notation and an inference engine for examples given in the work by Collins et al.
  - See: https://www.w3.org/Data/demos/chunks/reasoning/
- An implementation is invaluable for testing understanding of previous work and for identifying challenges for new work
- Collins distinguishes four kinds of plausible assertions
  - properties, relationships, implications and dependencies
- Inference involves qualitative parameters
  - certainty, typicality, similarity, frequency, dominance, conditional likelihood

Plausible Knowledge Notation (PKN)

- Properties

```
flowers of England |= daffodils, roses
(certainty high)
```

  - where |= means *includes*
  - and *certainty* is an example of a qualitative parameter

- Relationships

```
robin kind-of songbird
duck similar-to goose for habitat
duck dissimilar-to goose for neck-length
```

  - where *for* lists properties that the relationship applies to

- Dependencies

```
climate depends-on latitude
pressure decreases-with altitude
```

  - Describes a coupling between a pair of properties

- Implications

```
temperature of ?place = warm &
rainfall of ?place = heavy
   implies grain of ?place |= rice
```

  - A form of *if-then* rules

- Parameters can be given with all four kinds of statements

Relationships, implications and dependencies can be used for inferences in both directions

Example Knowledge Graph

a simple taxonomy
daffodils kind-of temperate-flowers
temperate-flowers kind-of flowers
flowers kind-of plants

used to infer that daffodils grow in England
flowers of England |= temperate-flowers
flowers of Netherlands |= daffodils, tulips
Netherlands similar-to England for flowers

used to infer climate of England
Netherlands similar-to England for climate
climate of Netherlands |= temperate

used to infer climate of Belgium
Belgium similar-to Netherlands for latitude
climate depends-on latitude

used to infer crop of Vietnam
climate of Vietnam = hot
rainfall of Vietnam = heavy

crop depends on climate and rainfall
climate of ?place = hot &
rainfall of ?place = heavy
  implies crop of ?place |= rice

The PKN syntax is up for discussion, e.g. using *includes*, *is*, and *is-not* in place of |=, = and != respectively?
The syntax will also be extended to support flexible quantifiers, quantification scope and related functions

Relationship to RDF and LPG

Resource Description Framework (RDF)

- RDF uses triples for labelled directed graph edges
- PKN relationships and properties can easily be mapped to triples
  - If you ignore value lists and parameter lists, but RDF-star would help …
- PKN statements correspond to sub-graphs in RDF
  - e.g. implication as a set of triples
- Considerations in respect to global names and ontologies
  - Role of lexicons for natural language

Labelled Property Graphs (LPG)

- LPG allows name/value pairs on graph edges and vertices
- LPG nodes correspond to collections of PKN properties
  - If you ignore parameter lists
- LPG links correspond to PKN relationships
  - Link properties expressed as PKN properties that cite the relationship
- LPG lacks implications and dependencies
  - these could be modelled as sub-graphs

Plausible reasoning is possible with RDF and LPG, but considerably harder than with PKN

Plausible Inferences

- Let’s start with something we want to find evidence for (or against*)
  - flowers of England |= daffodils
- We first check if this is a known fact and if not look for other ways to gather evidence
- We can either generalise the property value (daffodils)
  - flowers of England |= ?flower
- We find a matching property statement
  - flowers of England |= temperate-flowers
- We then look for ways to relate daffodils to temperate flowers
  - daffodils kind-of temperate-flowers
- This allows us to infer that daffodils grow in England
- Or we can either generalise the property argument (England)
  - flowers of ?place |= daffodils
- We look for ways to relate England to a similar country
  - Netherlands similar-to England for flowers
- We then find a related property statement
  - flowers of Netherlands |= daffodils, tulips
- This also allows us to infer that daffodils grow in England
  - The certainty depends on the parameters
- These examples use properties and relationships, but we can also look for implications and dependencies
  - e.g. a medium latitude implies a temperate climate, which in turn implies temperate flowers

* e.g. flowers of England != daffodils

Some challenges for study

- Quantifiers in queries and rules
  - Natural language is very flexible, e.g. *few* and *many*
  - Definite and indefinite references
  - Scoping to real or imagined contexts
- Example: are yellow roses found in England?

flower of England |= ?rose &
    ?rose kind-of rose &
    colour of ?rose |= yellow

- What about
  - Are all English roses red or white?
  - Are only a few roses yellow?
  - Are most people older than 20?

- yes/no answer to queries is often inappropriate
- Long lists are likewise awkward, and it is better to just give a few pertinent examples – but how to select them from the longer list?
- The computer should be prepared to respond by asking questions that clarify what's wanted
- Role of natural language dialogue
  - And asking for advice, explanations, etc.
- Grice's maxims of conversation
  - quantity, quality, relation and manner

Also want to explore supporting additional operators to = and |= as suggested by Burstein et al.

A Graph of Overlapping Graphs

- Very large knowledge graphs are difficult to deal with
- Displays of graphs lack context when zoomed in, and are intimidatingly complex when zoomed out
- This is also a challenge for automated reasoning due to the lack of context
- Proposed solution is to map large graphs into overlapping smaller graphs that model sets of *named contexts*
- For digital transformation of an enterprise, use subgraphs for different business functions
  - Different views for different departments
  - Dependency tracking for old & new uses
- Well defined business process for managing vocabularies at different levels of maturity
- Also pertinent for natural language

  "John opened the bottle and poured the wine"
  - i.e. a social situation with wine being transferred from bottle to guest's glasses
- Want to exploit *spreading activation* to identify shared contexts and most plausible word senses*
  - Activation wave weakened by high fan-out
- Specify contexts as named sets of PKN statements, analogous to sets of triples
  - Provides better scaling compared to defining contexts as sets of concepts

* Collins et al. suggest spreading activation could play a role for guiding search for potential inferences

Further Work

- Plan to develop a suite of related demos to cover plausible reasoning in respect to induction, abduction, planning, belief revision, causal, social and other kinds of reasoning, including embracing fuzzy reasoning and qualitative reasoning
- Implementing two ways to support distributed knowledge:
  - Cognitive agents with shared access to remote cognitive databases, analogous to different lobes in the cerebral cortex – hive minds with a shared memory and communal learning
  - Web of collaborating agents and humans for ecosystems of services – relation to *data spaces*
- Potential for single framework encompassing plausible reasoning & Bayesian Inference
  - Exploiting mix of qualitative and quantitive metadata for better inferences
- Research on indexing, metacognition, and System 1 & 2 style cognition*
- Work on integration with natural language interaction and natural language semantics
- Work on developing common sense skills through teaching and learning through interaction in real or virtual environments – once trained, agents are easily cloned
- Commercialisation in respect to Digital Transformation and enabling non-programmers to work with information using collaborative multi-modal cognitive agents

* See chunks & rules specification and demo's for the W3C Cognitive AI Community Group

“Machines that can use facts to present a convincing case could transform the way we make decisions – and help us understand our own rhetoric”

New Scientist, 7 September 2016 [Link]

Demo and Discussion

Contact: Dave Raggett dsr@w3.org

Example – does England have a temperate climate?

- No direct evidence found
- Look for indirect evidence that England has a temperate climate
  - Find England is part of Europe and look for evidence that Europe includes a temperate climate. It does, so it is likely that England has a temperate climate
  - Find that climate depends on latitude, and discover that a temperate climate is implied if the latitude is medium. Further find that England has a medium latitude, so it is likely that England has a temperate climate
  - Given that climate depends on latitude and that the Netherlands is similar to England in respect to latitude, and that the Netherlands has a temperate climate, so it likely that England does too.
- Look for indirect evidence that England doesn't have a temperate climate
  - None found

Multiple arguments in favour and none against, thus judge that it is true with high certainty

Example – does coffee grow in Llanos?

- No direct evidence found
- Look for indirect evidence that coffee grows in Llanos
  - Crop depends on climate and vegetation, and Llanos and Sao Paulo match on climate and vegetation, and coffee is grown in Sao Paulo, so conclude coffee is grown with medium certainty
- Look for indirect evidence that coffee doesn't grow in Llanos
  - Coffee implies high rainfall, but rainfall is medium and subject to a closed range, so conclude coffee is not grown with medium certainty

Plausible evidence for and against the premise, so no judgement is possible