

Ontology Summit 2019: Commonsense Explanations

Gary Berg-Cross and others TBD

1 Introduction

Commonsense reasoning and knowledge (CSK) was prominently featured as an early part of how Artificial Intelligence (AI) was conceptualized. It was assumed to be important in the development and enhancement of human-like, intelligent systems that include the ability to explain their reasoning and what they know. Commonsense research, since its inception, has focused on studying the consensus reality, knowledge, causality, and rationales that are ubiquitous in everyday thinking, speaking and perceiving as part of ordinary human interaction with the world. And as the name suggests, this knowledge and reasoning is common and available to the overwhelming majority of people, and manifest in human children's behavior by the age of 3 or 5. Commonsense and the ability to get along in an ordinary world was assumed as part of the original Turing Test, for example as discussed by Michael Grüninger. [8]

Some examples of CSK include the following type of reasoning:

- Taxonomic: Cats are mammals.
- Causality: Eating a pear makes you less hungry.
- Goals: I don't want to get hot so let's find shade.
- Spatial: You often find a toaster in the kitchen.
- Functional: You can sit on park benches if tired.
- Planning: To check today's weather look in a paper .
- Linguistic: The word "can't" is the same as "can not".
- Semantic: Dog and canine have a similar meaning.

More recently with Machine Learning (ML) capabilities of describing images, a visual Turing Test might involve question-answering based on real-world images, such as detecting and localizing instances of objects and relationships among the objects in a scene. Some questions that might be answered are the following:

- What is Person1 carrying?
- What is Person2 placing on the table?
- Is Person3 standing behind a desk?
- Is Vehicle 1 moving?

This article summarizes some of the research ideas discussed in the commonsense reasoning and knowledge track as part of the Ontology Summit 2019 on Explanations. [1] While both CSK and explanation are

distinct topics of current interest in AI and ML, they are related. The current emphasis on explanation, for example, grows in part out of the opacity of Deep Learning (DL) solutions for such tasks as labeling images or translating text (as discussed in more detail in other tracks of the Summit). These efforts motivate some opportunities for related work on commonsense that may be supportive. Of note this session on CSK examined issues around both commonsense and explanations particularly as they have been developing under the influence of modern ML and DL models. In part, the current excitement around ML grows out of the impact of big data and the recognition that one can now automate activities that previously required much a great deal of manual effort.

Compounding the challenge of explainability is the fact that current, rapid advances in AI include a great many approaches. Among these is the interestingly named Long short-term memory (LSTM) which is a particular artificial recurrent neural network (RNN) architecture (where output from previous steps are fed as input to the current step) used in the field of deep learning. These were developed to deal with the exploding and vanishing gradient problems that can be encountered when training traditional RNNs. LSTMs have proven a useful neural network (NN) architecture for processing sequential data such as speech/discourse or a series of visual images. Such applications, however, raise questions of what such systems know and what they can say about their knowledge and judgments.

2 Some History and Background to Commonsense and Explanation in AI

There is a long history showing the relevance of commonsense knowledge and reasoning to explanation. Certainly AI founders, such as John McCarthy, believed so and argued that a major long-term goal of AI should include endowing computers with standard commonsense reasoning capabilities. In “Programs with Common Sense” [16] described 3 ways for AI to proceed:

1. imitate the human central nervous system,
2. study human cognition or
3. “understand the common sense world in which people achieve their goals.”

Another related goal has been to endow AI systems with natural language (NL) understanding and production. It is easy to see that a system with both CSK and NL facilities would be able to provide smart advice as well as explanation of this advice. We see the relation of this in the early conceptualization of a smart advice taker system from McCarthy’s work that would have causal knowledge available to it for: “a fairly wide class of immediate logical consequences of anything it is told and its previous knowledge.” McCarthy further noted that this useful property, if designed well, would be expected to have much in common with what makes us describe certain humans as “having common sense.” He went on to use this idea to define CSK - “We shall therefore say that a program has common sense if it automatically deduces for itself a sufficiently wide class of immediate consequences of anything it is told and what it already knows.” [16]

In practice, back in the 70s and 80s, AI systems were not as the founders envisioned, They were brittle with handcrafted production rules that encoded useful information about diseases, for example. Consequently, this rule knowledge was fragmented and opaque and would break down in the face of obvious errors due in

part to a lack of common sense. It also came with a very simple, technical but not commonsense idea, of what was called an “explanation facility.” This was a technical/proof trace of rule firings which was thought to provide an explanation. Proofs found by Automated Theorem Provers can provide a map from inputs to outputs. Indeed in a narrow, logical sense the “Gold Standard” concept of explanation is deductive proof done using a formal knowledge representation (KR). Indeed there are multiple formalizations of deductive models involving some form of knowledge and reasoning. But as made clear in other summit sessions there are other forms, styles or meanings of explanation, and even mathematical proofs are not as reliable as one is often led to believe. One style of explanation concerns the provenance or source of some fact or statement. For example, “fact sentence F41 was drawn from document D73, at URL U84, in section 22, line 18.” This makes clear what the documented source of data is. That kind of explanation is just as important and allows follow up. Another type of explanation is the transparency of the algorithmic inference process - is it using sub-classing or something more complex like abduction? In some cases, it is very hard for a human to understand how the explanation was arrived at since the underlying structure is hidden. Other structuring of inference in presentation is possible using a clausal resolution graph or a Bayes net graph.

But an important question to ask is “do these make something clear?” They may provide the “how” an answer was arrived at in steps and which rules were involved, but not the justifying “why” of a satisfactory explanation. If a tree structure is involved in an explanation process we might get more of a “why” understanding with the possibility of drilling down and browsing the tree, having a focal point of attention on critical information or have the option of displaying a graphic representation that a human can understand. An example provided by Niket Tandon concerns a vehicle controller AI system explanation of driving based on visual sensing. [4] The system describes itself as “moving forward” as an activity while a human description is the more functional and common one of “driving down the street.” As to explanations the system says, “because there are no other cars in my lane” while the human explanation is “because the lane is clear.” These are similar but “clear” is a more comprehensive idea of a situation which might include construction, trees etc. A more elaborate example offered in how a smart system explains judgments of “what is a healthy meal.”

Commonsense assumptions and presumptions in a knowledge representation may be an important aspect of explanations and serve as a focus point. Ability to focus on relevant points may be part of the way a system is judged competent as is it’s perceived correctness - that it provides a good answer and a good explanation. Involved in such a judgment may also be evaluation of ethicality, fairness, and, where relevant, legality and various roles involved such as:

- Relational role,
- Processual role, and
- Social role.

An example of this is that the role of legal advice is different in the context of a banking activity compared to that of lying under oath. Part of the reason for limited explanations if not brittleness, was discovered by Clancey who found that Mycin’s individual rules play different roles, have different kinds of justifications, and are constructed using different rationales for the ordering and choice of premise clauses. [2] Since this knowledge isn’t made explicit it can’t be used as part of explanations. And there are structural and strategic concepts which lie outside early AI system rule representations. These can only be supported by appealing to some deeper level of (background) knowledge. One solution approach taken was to use ontologies to capture and formalize this knowledge. This made the argument that ontologies are needed

to make explicit structural, strategic and support knowledge which enhanced the ability to understand and modify the system (e.g. knowledge debugging as part of knowledge base (KB) development) as well as support suitable explanations.

To some extent efforts like CYC which started up in the 80s was an effort to avoid these problems by providing a degree of commonsense and modular knowledge. [14] CYC can provide a partial inference chain constructed in response to queries such as: “Can the Earth run a marathon?” In terms of a commonsense explanation we have a “no” because the Earth is not animate and the role capability of running a marathon is detailed by the knowledge in a sports module (called a micro-theory or MT). Around this time the issue for the development and application of ontologies was that the commonsense context was seldom “explicitly stated” and is difficult to do. But the need a formal mechanism for specifying a commonsense context had become recognized and some approach to it, such as CYC’ microtheories arose. These descended from J. McCarthy’s tradition of treating contexts as formal objects over which one can quantify and express first-order properties.

In the 80s CYC-type knowledge was also seen as important to associate systems which was arguing that “systems should not only handle tasks automatically, but also actively anticipate the need to perform them....agents are actively trying to classify the user’s activities, predict useful sub-tasks and expected future tasks, and, proactively, perform those tasks or at least the sub-tasks that can be performed automatically.” [13]

Around this time another influential CSK development was Pat Hayes’ “Naive Physics Manifesto” (1978) which proposed to develop a formal theory encompassing the large body of knowledge of the physical world. The vision was of an axiomatically dense KB using a unified conceptualization whose axioms were represented in symbolic logic. Emphasis was on formalizing foundational common “everyday” physical concepts, including: measurements and scales; spatial concepts (shape, orientation, direction, containment); substances and states (solids, liquids), physical forces, energy and movement; manufactured objects and assemblies. Much ontological work has followed the spirit of this idea if not the exact program outlined.

More recent work, reflecting the ability of ML systems to learn about visual information and even text has lead to more distinctions being made about computer science knowledge. An example from Tandon et al distinguished the visual modality expressing a type of knowledge (is it a simply observed property like color or some simple relation like part) from implications (shiny things imply smoothness and so less friction). [17, 21]

The KB would be supported by commonsense reasoners which might include answering “why” questions in an understandable way.

3 Additional Concepts of Explanations and Commonsense Understandings

It is worth noting in passing that CSK has been discussed in prior Ontology Summits. In particular the 2017 Summit on AI, Learning, Reasoning, and Ontologies and its track on “Using Automation and Machine Learning to Extract Knowledge and Improve Ontologies” is relevant but not all of it will not be discussed here. One prominent example from 2017 was Never-Ending Language Learner (NELL). Central to the NELL effort is the idea that we will never truly understand machine or human learning until we

can build computer programs that shares some similarity to the way human learn. In particular such systems, like people learn:

- many different types of everyday knowledge or functions and thus many contexts,
- from years of diverse, mostly self-supervised experience,
- in a staged curricular fashion, where previously learned knowledge in one context enables learning further types of knowledge,

using self-reflection and the ability to formulate new representations and new learning tasks enable the learner to avoid stagnation and performance plateaus.

As reported at the Ontology Summit 2017, NELL has been learning to read the web 24 hours/day since January 2010, and so far has acquired a knowledge base with over 80 million confidence weighted beliefs (e.g., `servedWith(tea, biscuits)`). [10] NELL has also learned millions of features and parameters that enable it to read these beliefs from the web. Additionally, it has learned to reason over these beliefs to infer (we might say using commonsense reasoning) new beliefs, and is able to extend its ontology by synthesizing new relational predicates. NELL learns to acquire two types of knowledge in a variety of ways. It learns free-form text patterns for extracting this knowledge from sentences on a large scale corpus of web sites. NELL also exploits a coupled process which learns text patterns corresponding type and relation assertions, and then applies them to extract new entities and relations. In practice it learns to extract this knowledge from semi-structured web data such as tables and lists. In the process it learns morphological regularities of instances of categories, and it learns probabilistic horn clause rules that enable it to infer new instances of relations from other relation instances that it has already learned. Reasoning is also applied for consistency checking and removing inconsistent axioms as in other knowledge graph (KG) generation efforts.

NELL might learn a number of facts from a sentence defining “icefield”, for example: “a mass of glacier ice; similar to an ice cap, and usually smaller and lacking a dome-like shape; somewhat controlled by terrain.” In the context of this sentence and this new “background knowledge” extracted it might then extract supporting facts/particulars from following sentences: “Kalstenius Icefield, located on Ellesmere Island, Canada, shows vast stretches of ice. The icefield produces multiple outlet glaciers that flow into a larger valley glacier.” It might also note not only the textual situation relating extracted facts but the physical location (e.g., Ellesmere Island) and any temporal situations expressed in these statements. A context that is important is that AI systems increasingly use advanced techniques such as deep learning which may in turn require some additional techniques to make them more understandable to humans and system designers as well as trusted.

Some factors making CSK (and explanation) more understandable (and hence trusted) were mentioned by Benjamin Grosz as part of his session presentation. [6] These include:

Influentiality which that heavily weighted hidden nodes and edges can affect discrimination of output, in a deep learning neural network (NN) or that some constructed approximation of an output, even if partial might provide a more understandable explanation. Such approximations may represent the top-3 weighted sub-models within some overall ML ensemble model. There is also the concept of “Model the model” wherein the learned NN model can then be learned as a secondary and simpler decision-tree model. Lateral relevance which involves interactivity for neighborhood, centrality, or top-k within categories exploration within graphs structures.

4 Contemporary Applications and Benefits of Explanation (that make sense)

There is an obvious benefit if semi- or fully-automatic explanations can be provided as part of decision support systems. This seems like a natural extension of some long used and understood techniques such as logical proofs. Benefits can be easily seen if rich and deep deductions could be supported in areas regarding policies and legal issues, but also as part of automated education and training, such as e-learning.

An example of this is the Digital Socrates application developed by Janine Bloomfield of Coherent Knowledge. The interactive tutor system will provide an answer to a particular topical question but also provides the logical chain of reasoning needed to arrive at the correct solution. Knowledge reuse and transfer is an important issue in making such systems scalable. Some Examples from Industry Using Explanation Technology among the examples offered by Benjamin Grosz [6] were:

- Coherent Knowledge - previously mentioned as providing eLearning systems with semantic logical deductive reasoning so proofs provide a natural deduction in programs using declarative, extended logic. But to make explanations understandable the systems employ NL generation along with a drill-down capability and interactive navigation of the knowledge space. Provenance of the knowledge is also provided.
- Tableau Software provides a different ability with specialized presentation of information via bar charts, etc.
- Kyndi - provides a more cognitive search in NL knowledge graphs (KGs). Here the capabilities include the focus on relevant knowledge including lateral relevance and provenance within an extended KG that is constructed using a combination of natural language processing (NLP), ML, and knowledge representation and reasoning (KRR).

A best practice architecture and functional example offered by Grosz et al [7] concerning Automated Decision Support for Financial Regulatory/Policy Compliance using Textual Rulelog software technology implemented with ErgoAI Suite. The system encodes regulations and related info as semantic rules and ontologies which support fully, robustly automate run-time decisions and related querying. Understandable full explanations are provided in English reflecting a digital audit trail, with provenance information. Rulelog helps handle the increasing complexity of real-world challenges such as found in data and system integration involving conflicting policies, special cases, and legal/business exceptions. For example it understands regulated prohibitions of banking transactions where a counterpart is an “affiliate” of the bank. Notably, Textual Rulelog (TR) extends Rulelog with natural language processing and uses logic to help do both text interpretation and text generation. With a proper use of rules mapping is much simpler and closer than with other KR’s and rulelog’s high expressiveness is much closer to NL’s conceptual abstraction level. As an added feature the system allowed what-if scenarios to analyze the impact of new regulations and policy changes.

5 Issues and Challenges Today in the Field of CSK (and Explanation)

The Issue of Making the Case for Modern ML and Deep Learning (DL) techniques have proven powerful for some problems such as in the field of computer vision such as employed for navigation or image identification. Research now reliable shows the value of transfer training/learning as part of this. So in pre-training a neural network model works on a known task, using stored images from a general source like ImageNet. The resulting trained neural network (with an implied “model”) used for new, but related, purpose-specific model. Of course it can be difficult to find training data for all types of scenarios and specific situations of interest. There are problems of representativeness and the seduction of the typical to some generalizations such “shiny surfaces typically hard, but some are not. There is the problem of perspective. The moon in the sky and a squirrel under a tree, which may be in the same image may seem the same size, but we know from experience that they are at different distances.

Many cognitive abilities developed in the first years of life provide the commonsense knowledge to handle these and problems like conservation of objects - if I put my toys in the drawer, they will still be there tomorrow. One may imagine handling such problems by building in commonsense knowledge or letting a system have training experience with conservation over time or place. These types of problems also arise with situational understanding - some important things are unseen but implied in a picture as part of the larger or implied situation such as an environmental or ecological one with many dependencies. An example offered by Niket Tandon was the implication of an arrow in a food web diagram which communicates “consumes” to a human (a frog consumes bugs). The problem for a deep net is that it is unlikely to have seen arrows used visually this way enough to generalize a “consumes” meaning. [4]

More recently related research has demonstrated that a similar “training” technique can be useful for many natural language tasks. An example is Bidirectional Encoder Representations from Transformers (BERT). So there is some general usefulness here but the black box nature and brittleness/distractability of the NN knowledge calls for some additional work. For example, as speaker Niket Tandon suggested, adversarial training (see [15]) during learning could help with mis-identification and commonsense considerations is one way to generate adversarial data.

There remain many problems with ordinary text understanding such as the implications and scope of negations and what is entailed. For example, Kang et al showed the problems of textual entailment with sentences from the Stanford Knowledge Language Inference set and how guided examples of an adversarial commonsense nature could help reduce errors to sentences like “The red box is in the blue box.” [12] And we can borrow from the previous ecological example as a language understanding example. A trained NN would not have seen insects, frogs and raccoons in one sentence frequently. To a human the use of an arrow as indicating consumption may be communicated in one trial situational learning sentence - “this is what we mean by an arrow”. So again the higher level of situations and situational questions may require commonsense, to understand the circumstances. Handling focus and scale is another problem in visual identification. In a lake scene with a duck a ML vision system may see water features like dark spots as objects. Here as Niket Tandon argued deep neural nets can easily be fooled (see [18]) so there may be a need for a model of the situation and what is the focus of attention - a duck object. Some use of commonsense as part of model-based explanations might help during model debugging and decision making to correct the apparently unreasonable predictions. [4]

In summary, Niket Tandon suggests that the above reasons, commonsense aware models and representa-

tions may be assumed to be DL friendly may provide several benefits especially for the NN and DL efforts underway. They may:

- help to create adversarial learning training data (for a DL)
- help to generalize to other novel situations, alternate scenarios and domains
- compensate for limited training data
- be amicable to and facilitate explanation e.g. with intermediate structures

In the sub-sections below these points are further illustrated by some reasons for the challenges to DL.

To understand situations (i.e., what exactly is happening?) a naive computational system has to track everything involved to a situation/event. This may involve a long series of events with many objects and agents. The ecological example provided before is illustrative as is visualizing a play in basketball even as simple as a made or missed dunk. Images of activities can be described by these three NL sentences: “He charges forward. And a great leap. He made a basket.” These sentences may be understood in terms of some underlying state-action-changes. There is a sequence of actions such as running and jumping, but there are also implied states as follows:

1. The ball is in his hands. (not actually said, but seen and important for the play)
2. The player is in the air. (implied by the leap)
3. The ball is in the hoop. (technically how a basket is made)

We can represent the location of things in the three sentences above like this:

1. Location (ball) = player’s hand
2. Location (player) = air
3. Location (ball) = hoop

The thing is that these all fit into a coherent action with the context of basketball, and we know and can focus on the fact that the location of the ball at the end of the jump is a key result. On other hand, as shown by Dalvi, Tandon, Clark, it is expensive to develop a large enough training set for such activities, and the resulting state-action-change models have so many possible inferred candidate structures (e.g., Is the ball still in his hand? Maybe it was destroyed.) that common events can evoke an NP-complete problem. [17, 21] Without sufficient data (remember it is costly to construct), the model can produce what one would consider to be absurd, unrealistic choices based on commonsense experience such as the player being in the hoop. Yet, they are physically possible nevertheless.

A solution is to have a commonsense aware system that constrains the search for plausible event sequences. This is possible with the design and application of a handful of universally applicable rules. For example these constraints seem reasonable based on commonsense:

- An entity must exist before it can be moved or destroyed. (certainly not likely in basketball)
- An entity cannot be created if it already exists.

In the work discussed by Niket and Tanden these constraints were directly derivable from the Suggested Upper Merged Ontology (SUMO) rules such as: MakingFn, DestructionFn, MotionFn. This provides preliminary evidence that ontologies, even early ones such as SUMO, could be good guides for producing a handful of generic hard constraints in new domains. So how much help do these constraints provide? Commonsense biased search improves precision by nearly 30% over State Of The Art DL efforts - Recurrent Entity Networks (EntNet) [9], Query Reduction Networks (QRN) [19], and ProGlobal. [17, 21]

Generally, the current state of the art for ML suggests that deep learning can provide some explanations of what they identify in simple visual datasets such as Visual Query Answering (VQA) and CLEVR. [5, 11] They can answer questions like “What is the man riding on?” in response to an image. Whereas, commonsense knowledge is more important when the visual compositions are more dynamic and involve multiple objects and agents.

5.1 How to acquire CSK

As noted earlier it can be costly but some ruling constraints can be derived from existing ontologies. An extraction process from text using AI, ML or NLP tools may be less costly but much noisier and has the problem of placing knowledge in situational context. In some advanced machine learning cases prior knowledge (background knowledge) may be used to judge the relevant facts of an extract, which makes this a bit of a bootstrapping situation. However much of what is needed may be implicit and inferred and is currently only available in unstructured and un-annotated forms such as free text. NELL is an example of how NLP and ML approaches can be used to build CSK and domain knowledge but source context as well as ontology context needs to be taken into account to move forward. It seems reasonable that the role of existing and emerging Semantic Web technologies and associated ontologies is central to making CSK population viable and that some extraction processes using a core of CSK may be a useful way of proceeding. Another problem is that there are known biases such as reporting bias involved in common human understanding. Adding to this is the knowledge exist in multi-modal and related forms which makes coverage challenging across different contextual situations. Some of these contextual issues were discussed as part of the Ontology Summit 2018.

5.2 Confusion about explanation and commonsense concepts

It is perhaps not surprising that there is confusion between explanation and commonsense concepts among non-research industry and the general public. Unfortunately, this confusion is also common in technical media that cover computation and systems, although usually not when covering AI. Like many specialties this needs to be addressed first in the research community where useful consensus can be reached and a common way of discussing these be developed and scaled down to be understood by others.

5.3 The Range of Ontologies Needed for an Adequate CSK

Michael Grüninger’s work suggested some significant types of ontologies might be needed to support something as reasonable as a Physical Embodied Turing Test. [8] The resulting suite is called PRAXIS (Perception, Reasoning, and Action across Intelligent Systems) with the following components:

- Solid Physical Objects (SoPhOs)
- Occupy (Location - Occupation is a relation between a physical body and an spatial region)
- Process Specification Language (PSL)
- Processes for Solid Physical Objects (ProSPerO)
- Ontologies for Video (OVid)
- Foundational Ontologies for Units of measure (FOUnt)

5.4 Mission creep

Expansivity of task/aspect or “mission creep” is a common phenomenon in AI among theorists, especially in hot, new areas such as Deep Learning. One sees the expansion of the topic, for example, at the IJCAI-18 workshop on explainable AI, which included a wide range of topical questions drawn from several disciplines, including cognitive science, human factors, and psycholinguistics, such as:

- How should explainable models be designed?
- How should user interfaces communicate decision making?
- What types of user interactions should be supported?
- How should explanation quality be measured?

A perennial problem is ignorance of past research and a sense of what’s already practical. For example, in the area of deep policy/legal deduction for decisions, is it difficult to provide a full explanation of extended logic programs, along with NL generation and interactive drill-down navigation? We seem to have reached a point where we can have a good version of cognitive search that takes into account provenance along with focus and lateral relevance, using extended knowledge graphs.

5.5 How to evaluate explanations or CSK

Evaluation remains a research issue in both CSK and explanation. As we have seen, there are many criteria for judging the goodness of automated processes from simple labeling to full explanations. How does one use such criteria to quantitatively measure the quality of CSK and explanation? In some cases regular domain knowledge will not support commonsense reasoning and some enhancement is needed. Micheal Grüninger cited the example of translating a domain ontology to a domain state ontology. [8] In this enhancement, activity occurrences correspond to mappings between models of the domain ontology. And we would need to classify activities with respect to possible changes. An implication here is that some additional ontological steps are needed for such an enhancement. As example, Grüninger offered an approach looking at fluent - verbs. For each verb, the process is to:

- Identify the fluents (i.e. states of the world, leveraging PSL and the situation calculus) that are changed when the process corresponding to the verb occurs.
- Axiomatize the domain state ontologies that contain the fluents in their signature
- Identify the domain ontologies for these domain state ontologies. This may require additional axiomatizations if such ontologies have not previously been considered.

- Axiomatize the domain process ontologies.
- The process corresponding to the verb will either be an atomic activity in one of the domain process ontologies or it is a complex activity that is composed of activities in the domain process ontologies.

It is not as simple as saying that a system provides an exact match of words to what a human might produce given the many ways that meaning may be expressed. And it is costly to test system generated explanations or even captions against human ones due to the human cost. One interesting research approach is to train a system to distinguish human and ML/DL system generated captions (for images etc.). After training one can use the resulting criteria to critique the quality of the ML/DL generated labels. [3] A particular task is evaluating the quality of knowledge, both CSK and non) extracted from text. Provenance information or source context is one quality needed. In some cases, and increasingly so, a variety of CSK/information extracted is aligned (e.g. some information converges from different sources) by means of an extant (hopefully quality) ontology and perhaps several. This means that some aspect of the knowledge in the ontologies provides an interpretive or validating activity when building artifacts like KGs. Knowledge graphs can also be filled in by internal processes looking for such things as consistency with common ideas as well as from external processes which adds information from human and/or automated sources. An example currently used is to employ something like Freebase’s data as a “gold standard” to evaluate data in DBpedia which in turn is used to populate a KG. We can again note that a key requirement for validatable quality of knowledge involves the ability to trace back from a KB to the original documents (such as LinkedData) and if filled in, from other sources such as humans to make it understandable or trustworthy. It is useful to note that this process of building such popular artifacts as KGs clearly shows that they are not equivalent in quality to supporting ontologies. In general there some confusing of equating the quality of extracted information from text, KGs, KBs, the inherent knowledge in DL systems and ontologies.

5.6 Alternative values and disconnects between users, funders and investors

Users often perceive critical benefits/requirements for things like commonsense explanation. Their context is that their current AI systems may produce confusing results with explanations that do not make sense and/or are complex and unwieldy, while often being misleading or inaccurate. Funders certainly consider big picture benefits such as more overall effectiveness of systems with less exposure to risk of non-compliance. But they should also look at research issues, such as agile development and knowledge maintenance involved and current methods for producing explanations and/or commonsense reasoning are expensive and perhaps not scalable or take too much time to update. Investors (both venture and enterprise-internal) may look for a shorter and more dramatic feature impact and business benefits compared to currently deployed methods. Among those benefits is accurate decisions but a case should be made for better explanations that can be cost effective by requiring less labor or the precious time of subject matter experts which now can participate in a closer knowledge refinement loop. Investors fail to perceive value of what they consider add-ons like explanation which seem costly, but especially commonsense reasoning which is seen as simple.

5.7 How to represent knowledge for explanations

Both formal/logical and informal or commonsense knowledge of assertions, queries, decisions, answers and conclusions along with their explanations has to be represented. As noted in the work of Grüniger single ontologies are not likely to be suitable as work expands and more contexts are encountered. This will require multiple ontologies. Big knowledge, with its heterogeneity and depth complexity, may be as much of a problem as Big Data especially if we are leveraging heterogeneous, noisy and conflicting data to create CSK and explanations. It is hard to imagine that this problem can be avoided. The ontology experience is that as a model of the real world we need to select some part of it based on interest and conceptualization of that interest. Differences of selection and interpretation are impossible to avoid and it can be expected that different external factors will generate different context for any intelligent agent doing the selections and interpretation needed as part of a domain explanation.

In the context of applying highly structured knowledge such as ontologies to DL system knowledge there is an issue of how to first connect the noisy space of data to the perfect world of ontologies/KBs and their modules, such as CYC microtheories. Such connections may not be scalable, and the effort not easily made crowdsourcable. [17, 21] Various approaches exist for different forms of CSK and the integration of these is challenging. Linked data may view some formal knowledge as a set of linked assertions but for integration these may be linked to regular sentences expressing these assertions from which sentences may be generated. While commonsense is an important asset for DL models, its logical representation such as microtheories has not been successfully employed. Instead, tuples or knowledge graphs comprising of natural language nodes has shown some promise, but these face the problem of string matching i.e., linguistic variations. More recent work on supplying commonsense in the form of adversarial examples, or in the form of unstructured paragraphs or sentences has been gaining attention recently. But experience in this area suggests that construction is noisy/approximate when scaling up to significant amounts of diverse data and that task or application specific KBs may be more effective if scale is involved.

In the previously cited example (Financial Regulatory/Policy Compliance), NL-syntax sentence may have one or more logic-syntax sentences associated with it that formally encode the understood assertions and allow some reasoning (unlike free text which can represent CSK). For explanations they may also assert its provenance, or even represent its text interpretation in some other expression (this means that). Obviously logic-syntax sentences may have one or more NL-syntax sentences that can be expressively associated with it which can be output by a text generation process. There may also be a round trip view in that there are source sentence used in text interpretation, that produced a logical representation from which other expression of the knowledge may be produced. Work at this level is still in its early stages.

5.8 Robustness and Understandability

How robust is commonsense reasoning and decision making? As noted there are now many ML applications which are increasingly looked on as mature enough to use for some ordinary tasks. Visual recognition is one of these but Shah et al suggests that some such applications are not robust. [20] Simple alternative NL syntactic formulations lead to different answers. For example, “What is in the basket” and “What is contained in the basket” (or “what can be seen inside the basket”) evoke different answers. Humans understand these as similar, commonsense meanings, but ML systems may have learned something different. Another problem is that DLs, by themselves, are black-box in nature. So while these approaches allow powerful predictions, by themselves their outputs cannot be directly explained. Additional functionality

is needed including a degree of commonsense reasoning. Such focused, good, fair explanations may use natural language understanding even be part of a conversational dialogue human-computer interaction (HCI) in which the system uses previous knowledge of user (audience) knowledge and goals to discuss output explanations. Such “Associate Systems” may get at satisfactory answers because they include a capability to adaptively learn user knowledge and goals and are accountable for doing so over time as is commonly true for human associates.

5.9 Enhancing Ontology Engineering Practices

We will need to arrive at a focuses understanding of CSK that can be incorporated into ontological engineering practices. For efforts like CSK base building this should include guidance and best practices for the extraction of rules from extant, quality ontologies. If knowledge is extracted from text and online information building of CSK will require methods to clean, refine and organize them probably with the assistance of tools. In light of this future work, will need to refine a suite of tools and technologies to make the lifecycle of CSKBes easier and faster to build.

6 Conclusions

It seems clear that both CSK and explanation remain important topics as part of AI research and its surging branch of ML. Further they are mutually supportive although explanation may be the more active area of diverse work just now. We seem close to AI systems that will do common tasks such as driving a car or give advice on common tasks like eating. It seems clear that such everyday tasks need to exhibit robust commonsense knowledge and reasoning to be trusted. Thus as intelligent agents become more autonomous, sophisticated, and prevalent, it becomes increasingly important that humans interact with them effectively to answer such questions as “why did my self-driving vehicle take an unfamiliar turn?” Current AI systems are good at recognizing objects, but can’t explain what it sees in ways understandable to laymen. Nor can it read a textbook and understand the questions in the back of the book which leads researchers to conclude it is devoid of common sense. We agree as DARPA’s Machine Common Sense (MCS) proposal put it that the lack of a common sense is “perhaps the most significant barrier” between the focus of AI applications today (such as previously discussed), and the human-like systems we dream of. And at least one of the areas that such an ability would play is with useful explanations. It may also be true, as NELL researchers argue, that we will never produce true NL understanding systems until we have systems that react to arbitrary sentences with “I knew that, or didn’t know and accept or disagree because X”.

Some General Recurring Questions we suggest are worth considering include:

1. How can we leverage the best of the two most common approaches to achieving commonsense? formal representations of commonsense knowledge (e.g. encoded in an ontology’s content as in Cyc or Pat Hayes’ Ontology of Liquids) vs. strategies for commonsense reasoning (e.g. default reasoning, prototypes, uncertainty quantification, etc.)
2. How to best inject commonsense knowledge into machine learning approaches? Some progress on learning using taxonomic labels, but just scratches the surface

3. How to bridge formal knowledge representations (formal concepts and relations as axiomatized in logic) and representations of language use (e.g. Wordnet)

Commonsense knowledge and reasoning could assist in addressing some challenges in DL as well as explanation and in turn interest in these can turn to CSK and reasoning for assistance.

Many challenges remain including those of adequate representative knowledge, how to acquire the proper knowledge and the number of different but consistent and supportive ontologies may be needed even for simple task. How to use these to mitigate the brittleness of DL models in light of various adversarial/confusing alternative inputs is a problem as is the brittleness explanations.. Among the problems that DL systems face as they address increasingly complex situations is producing generalization as humans seem to do for unseen situations when faced with limited training data/experience. Another problem is that DL systems are not aware of an overall context when processing subtle patterns such as social interactions. A recent commonsense aware DL model makes more sensible predictions despite limited training data. Here, commonsense is the prior knowledge of state-changes e.g., it is unlikely that a ball gets destroyed in a basketball game scenario. Commonsense knowledge and reasoning can compensate for limited training data and make it easier to generate explanations, given that the commonsense is available in an easily consumable representation.

The model injects commonsense at decoding phase by re-scoring the search space such that the probability mass is driven away from unlikely situations. This results in much better performance as a result of adding commonsense.

References

- [1] K. Baclawski, editor. *Ontology Summit 2019: Explanations*, 2019. Retrieved on April 28, 2019 from <http://bit.ly/2z0JGY4>.
- [2] W. Clancey. Explainable ai past, present, and future: A scientific modeling approach, February 2019. Retrieved on April 28, 2019 from <http://bit.ly/2Scjvo6>.
- [3] Y. Cui, G. Yang, A. Veit, X. Huang, and S. Belongie. Learning to evaluate image captioning. *ArXiv*, 2018. arXiv:1806.06422.
- [4] G. deMelo, N. Tandon, and A.S. Varde. Commonsense for machine intelligence: Text to knowledge and knowledge to text, Nov 2017.
- [5] Y. Goyal, T. Khot, D. Summers-Stay, D. Batra, and D. Parikh. Making the V in VQA matter: Elevating the role of image understanding in Visual Question Answering. In *Conference on Computer Vision and Pattern Recognition (CVPR)*, 2017. Retrieved on April 28, 2019 from <https://visualqa.org/>.
- [6] B. Grosz. An overview of explanation: Concepts, uses, and issues, January 2019. Retrieved on April 28, 2019 from <http://bit.ly/2R9iD30>.
- [7] B. Grosz, J. Bloomfield, P. Fodor, M. Kifer, I. Grosz, M. Calejo, and T. Swift. Automated decision support for financial regulatory/policy compliance, using textual rulelog. *Financial Times*, 2014.

- [8] M. Grüninger. Ontologies for the physical turing test, January 2019. Retrieved on April 28, 2019 from <http://bit.ly/2RiQSFz>.
- [9] M. Henaff, J. Weston, A. Szlam, A. Bordes, and Y. LeCun. Tracking the world state with recurrent entity networks. 2017. Retrieved on April 28, 2019 from <https://arxiv.org/pdf/1612.03969>.
- [10] E. Hruschka. Never-Ending Learning approach for populating and extending an ontology. In *Ontology Summit 2017: AI, Learning, Reasoning, and Ontologies*, March 8 2017. Retrieved December 9, 2017 from <http://bit.ly/2xYeEjN>.
- [11] J. Johnson, L. Fei-Fei, B. Hariharan, C.L. Zitnick, L. van der Maaten, and R. Girshick. CLEVR: A diagnostic dataset for compositional language and elementary visual reasoning. *ArXiv*, 2016. Retrieved on April 28, 2019 from <https://arxiv.org/pdf/1612.06890.pdf>.
- [12] D. Kang, T. Khot, A. Sabharwal, and E. Hovy. AdvEntuRe: Adversarial training for textual entailment with knowledge-guided examples. In *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Long Papers)*, pages 2418–2428, Melbourne, Australia, July 2018. Association for Computational Linguistics.
- [13] A. Kobsa and W. Wahlster, editors. *User Models and Dialog Systems*. Springer-Verlag, 1989.
- [14] D. Lenat. CYC: A large-scale investment in knowledge infrastructure. *Communications of the ACM*, 38(11):33–38, Nov. 1995. Retrieved on April 28, 2019 from <http://dx.doi.org/10.1145/219717.219745>.
- [15] Y. Li, T. Baldwin, and T. Cohn. What’s in a domain? learning domain-robust text representations using adversarial training. *ArXiv*, 2018. arXiv:1805.06088.
- [16] J. McCarthy. Programs with common sense. In M. Minsky, editor, *Semantic Information Processing*, pages 403–418. MIT Press, 1968. Originally published in 1959.
- [17] B.D. Mishra, L. Huang, N. Tandon, W. Yih, and P. Clark. Tracking state changes in procedural text: A challenge dataset and models for process paragraph comprehension. In *Proc. NAACL*, 2018.
- [18] A. Nguyen, J. Yosinski, and J. Clune. Deep neural networks are easily fooled: High confidence predictions for unrecognizable images. In *Computer Vision and Pattern Recognition (CVPR ’15)*. IEEE, 2015.
- [19] M. Seo, S. Min, H. Hajishirzi, and A. Farhadi. Query-reduction networks for question answering. 2017. Retrieved on April 28, 2019 from <https://arxiv.org/pdf/1606.04582>.
- [20] M. Shah, X. Chen, M. Rohrbach, and D. Parikh. Cycle-consistency for robust visual question answering. *ArXiv*, 2019. arXiv:1902.05660.
- [21] N. Tandon, B.D. Mishra, J. Grus, W. Yih, A. Bosselut, and P. Clark. Reasoning about actions and state changes by injecting commonsense knowledge. In *Proc. EMNLP*, 2018.