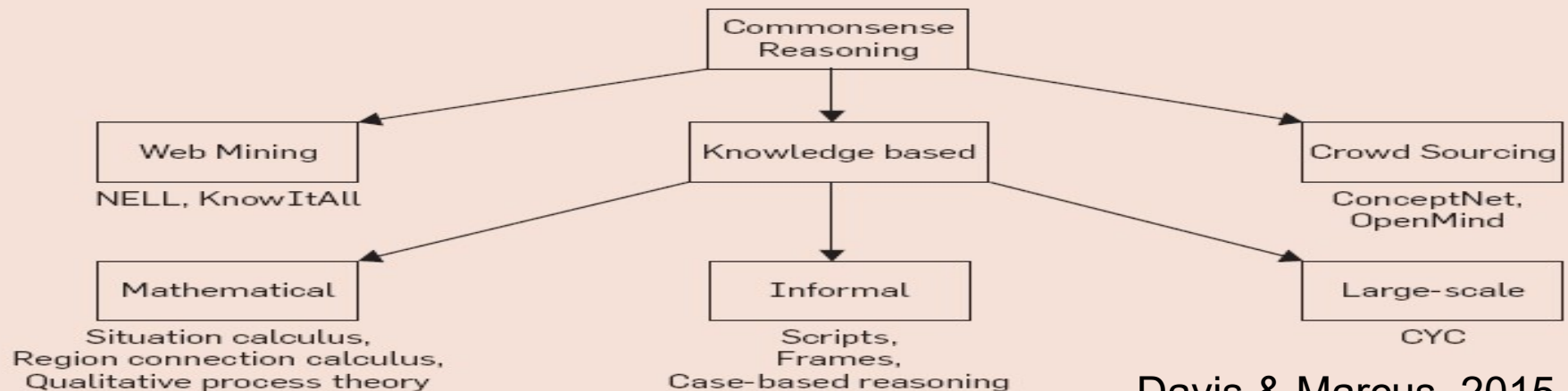


AI Explanations and Commonsense Knowledge (CSK)

Co-Champions: Gary Berg-Cross (Consultant) &
Torsten Hahmann (U of Maine)

Commonsense Knowledge & Reasoning along with Explanation
are ubiquitous in
everyday thinking, speaking and perceiving
as part of our ordinary interaction with the world



Davis & Marcus, 2015

Topical Outline

- The following is a preliminary summary of material from the Commonsense Track which may also be useful for the Summit Communique.
- Introduction/Background
 - Examples of CSK
 - Additional Views of Explanation
 - A brief History of Commonsense Knowledge and Relations Explanations in AI.
- Summary using slides from the speaker presentations
 - Gruninger, Grosz and Tannen
- Cross-Track Integration
- Opportunities, Challenges and issues to investigate
- Preliminary Findings and Conclusions.

What is CSK? Examples

- 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.
- Examples of commonsense and the ability to get along in an ordinary world:
 - Taxonomic: Cats are mammal 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: "can't" is the same as “can not” Semantic: dog and canine have a similar meaning
- Such knowledge was assumed the original Turing Test and as discussed in Session 1 by Michael Gruninger It might involve question-answering based on real-world images as part of the Visual Turing Test such as detecting and localizing instances of objects and relationships among the objects in a scene.
- What is Person1 carrying? What is Person2 placing on the table?
- Is Person3 standing behind a desk Is Vehicle 1 moving?
- 5/7/19 CSK Track for 2019 Ontology Summit Symposium

Brief History of Commonsense Knowledge and Reasoning & Explanations in AI.

- Commonsense reasoning and knowledge (CSK) was prominently featured as an early part of how Artificial Intelligence (AI) was conceptualized (McCarthy).
- It was assumed to be important in development and enhancement of human-like, intelligent systems including the ability of system 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.
- Early rule based systems and explanation as rule tracing proved brittle & less informative about the why of a decision.
- More explicit knowledge such as in ontologies became a direction.
 - CYC effort to formalize background knowledge.
 - Hayes Naive Physics Manifesto” (1978) which proposed to develop a formal theory encompassing the large body of knowledge of the physical world.
- Associate systems and NLP advances were notable drivers of both CSK & explanation.

More Recent AI & ML Efforts Provide Additional Factors Involved in Explanation

- 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,
- Some factors making CSK (and explanation) more understandable (and hence trusted) were mentioned by Benjamin Grosz as part of his session presentation. These include :
- Influentiality which that heavily weighted hidden nodes and edges can affect discrimination of output, in a deep learning (DL) neural network (NN) or that some constructed approximation of an output, even if partial might provide a more understandable explanation.
 - Example - top-3 weighted sub-models within some overall ML ensemble model.
- “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.
- The role of NL 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 need as is Visualization, Transparency (of an algorithm/process for inference) and drill down.

Summary from our speaker's presentations:
Gruninger (Physical Embodied Turing Test), Grosz and Tanden

- Michael Gruninger's work gets at some of significant types of ontologies and knowledge needed for CSK & explanation.
- Ontology Suite is called PRAXIS (Perception, Reasoning, and Action across Intelligent Systems) with the following components:
 - SoPhOs (Solid Physical Objects)
 - Occupy (location - Occupation is a relation between a physical body and an spatial region - there is no mereotopological relationship between spatial regions and physical bodies)
 - PSL (Process Specification Language)
 - ProSPerO (Processes for Solid Physical Objects)
 - OVID (Ontologies for Video)
 - FOUnt (Foundational Ontologies for Units of measure)

Summary from Benjamin Grosz Presentation – many concepts involved in Explanation (as well as CSK) and many are Issues

- Is it a Summarization, grain (coarse vs. fine), drill-down, elaboration?
- Or Partial vs. complete? Or approximate rather than precise
- Is it a structuring of inference in a presentation (E.g., a clausal resolution graph or a Bayes net graph)
- Are assumptions and presumptions made explicit or can be queried?
- Does it target and include user knowledge and goals (i.e., commonsense or fits user model)
- Is natural language (NL) or graphical presentation generated?
- What is the source of terminology used?– e.g., ontologies, thesauri, vocabularies

- **Many Issues**

- Are there disconnects among researchers, industry and media, or between users and investors with respect to what constitutes an acceptable or successful explanation?

Grosof Case Study: Automated Decision Support for Financial Regulatory/Policy Compliance

Problem: Current methods are expensive and unwieldy, often inaccurate

Solution Approach – using Textual Rulelog software technology:

Textual Rulelog (TR) extends Rulelog with NLP 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.

- *Encode regulations and related info as semantic rules and ontologies*
- Fully, robustly automate run-time decisions and related querying
- Provide understandable full explanations in English
 - *Proof*: Electronic audit trail, with provenance
- Handles increasing complexity of real-world challenges
 - Data integration, system integration
 - Conflicting policies, special cases, exceptions
 - What-if scenarios to analyze impact of new regulations and policies

Benefits – compared to currently deployed methods:

- More Accurate and **Better explanations**
- More Cost Effective & More Agile – faster to update
- More Overall Effectiveness: less exposure to risk of non-compliance

ErgoAI Architecture (Grososf)

Optional Custom Solutions

ErgoAI suite

ErgoAI Studio
(Integrated Development Env.)
GUI for explanation, editing,
querying, debugging, tuning

ErgoAI Reasoner
querying, compilation, cacheing,
explanation generation,
execution monitoring

Knowledge Base

API's
Java, C
Python
WS

API's
Java, C
Python
WS

Users

External Services &
Frameworks

Relational DB

RDF/Graph DB

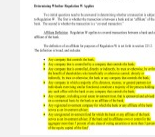
Other Sem. Tech

Machine Learning

Apps, Docker, ...

Complex Information

- English/NL Doc.'s etc.
- Policies, Regulations
- Financial, Legal,
Medical, Science



External Info
(multi-source)

- Data
- Views, **Rules**
- **Schemas & Ontologies**
- **Results of ML**

App
Actions

events,
decisions

WS = Web Services. Sem. = Semantic. ML = Machine Learning

© Copyright Coherent Knowledge Systems, LLC, 2018. All Rights Reserved.

Niket Tanden's (Visual/Word) ML/DL Challenges (that CSK helps with)

- **Robustness (needed for Visual Question Answering)**
- **Deep neural networks are easily fooled: evolvingai.org/fooling**
- **Get stuck “flirting with words”**
- **The need for adversarial training an/or CSK**
- **Difficult to find training data for all types of scenarios (we use context)**
- **Unforeseen circumstances in simple visual models (what does an → imply?)**
 - **Situational questions require commonsense, to model the situation.**
 - **Chaining makes for a good explanation, but linking facts would benefit from commonsense.**

Niket Tanden on DL, Explanations, and CSK

- For a machine to **understand** what exactly is happening, it has to track everything involved (Basketball dunk example). Location (player) = air
- DL friendly commonsense representations are useful for both explanation and compensate for weaknesses in DL.
- Recent work, reflecting the ability of (multimodal) ML systems to learn about visual information and even text has lead to more needed distinctions being made about CS knowledge.

Example, Tanden et al (2018) 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).

Properties of agents (emotions) in difficult circumstances, while commonly known, are more implicit as are the actions involved in fixing a tire or avoiding a traffic jam.



5/7/19

<i>Q: Is this a healthy meal?</i>	Textual Justification	Visual Pointing
	 A: No	
	 A: Yes	

Cross-Track Integration

- As Speaker Sargur Srihari noted Probabilistic Approach are important in AI Explanations.
 - Traditionally knowledge being represented using probabilistic graphical models such as Bayesian networks, Markov networks, Restricted Boltzmann machines, etc.
 - More recent efforts are better suited for explainability , which has methods for most probably explanation (MPE).
- Interpretable models using ideas such as causality are important as is the approach to induce “models” from DL systems.
- In XAI there is also the idea of reasonableness/CSK of explanation.
- Handling Financial system requirements using ErgoAI which was implementation of Rulelog (Grosz) and featured:
 - Semantic ncode regulations and related info allows fully, robustly automate run-time decisions and related querying
 - Handles increasing complexity of real-world challenges
 - Provide understandable full explanations in English
 - *Proof is an* Electronic audit trail, with provenance.

Opportunities (Tanden)

- CSK and explanation can complement one another
- (But a suite of CSK ontologies may be needed)
- CSK can be particularly valuable for ML/DL efforts:
- Commonsense aware models may:
 - help to create adversarial learning training data
 - generalize to other novel situations and domains
 - compensate for limited training data
 - be amicable to explanation e.g. with intermediate structures.

Some Challenges and Issues to Investigate

- Confusion about what is meant by explanation and commonsense concepts & how their quality should be measured.

How to clarify DL & ML “black box” knowledge to provide acceptable explanations.

- Disconnents (between users, funders and investors) -there is also general ignorance about what is already possible vs. what is well beyond current capabilities.
- How to acquire CSK efficiently, controlling mission creep, so computer-based explanations can be made as informative and useful as an explanation from a well-informed person?
- How to represent knowledge for explanations including integrating ontological and statistical approaches?
- Understanding the quality & variety of ontologies, such as addressing CSK, to support explanation and constrain DL....

Preliminary Findings/Conclusions - CSK is helpful for both explanation and as a complement to AI and DL efforts.

- AI will do` common tasks - driving a car or giving advice on common tasks (eating).
- Such everyday tasks need to exhibit robust commonsense knowledge and reasoning to be trusted.
- Autonomous, sophisticated, and prevalent, intelligent agents need to interact with humans 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/users.
- 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.*
- .