Ontology Summit 2017
Track A Wrapup:

The mission of this track was to scope out challenges, opportunities, current & emerging practices in "Using Automation and Machine Learning to Extract Knowledge and Improve Ontologies"

Champion: Gary Berg Cross
May 15, 2017
Outline

1. Problems/Vision/Opportunities:
   1. Use Automation to Overcome Knowledge Bottleneck for Quality Knowledge/Ontologies

2. Our Speakers & Session Topics

3. Knowledge Extraction Challenges

4. Relations (challenges & opportunities) across tracks
   1. Big data & IoT, hybrids, more cognitive systems etc...

5. Challenging Future Prospects - research & practical issues
Why Automation? Motivation But Knowledge Bottleneck

- Big Science, Data, Government, Industry & IoT provide many motivating challenges.
  - Text to structured Information
  - Heterogeneity & context resistant to simple processing
  - Need glue is needed to integrate & harmonize knowledge.
Opportunity of New AI/ML?

- Early AI was optimistic of super intelligence in 20\textsuperscript{th} century (J Sowa)
- But there remains a knowledge bottleneck
  - Cyc (1984 on) can it update its KB & learn by reading a textbook?
  - But there is good progress in NLP and ML -NELL can “read”
    - Progress in ontological knowledge to help these?
- And now we have deeper, multi-layer Neural Nets
  - Used to learn time-varying patterns (images, sounds, games)
  - More may be needed via un/semi-supervised learning
    - Multi-sources of info and hybrid architectures
    - Need more than NLP, need NLU
Overview & Session 1 Speakers

Launch
1. Gary Berg-Cross Overview of the topic
   Approaches to helpful automation dealing with the knowledge bottleneck.

Overview

Session 1 (March)
3. Estevam Hruschka “Never-Ending Language Learning (NELL)”
5. Alessandro Oltramari, "From machines that learn to machines that know: the role of ontologies in machine intelligence"
Our Three Session 2 Speakers

Session 2 (April)

1. Michael Yu (UCSD)
   "Inferring the hierarchical structure and function of a cell from millions of biological measurements".

2. Francesco Corcoglioniti (Post-doc at Fondazione Bruno Kessler, Italy)
   "Frame-based Ontology Population from text with PIKES"

3. Evangelos Pafilis (Hellenic Center Marine Research [HCMR]) -
   “EXTRACT 2.0: interactive extraction of environmental and biomedical contextual information."
Ontology Learning Opportunity?    Paul Buitelaar
Insight Centre for Data Analytics, National University of Ireland, Galway

\[ \forall x \ ( \text{country}(x) \rightarrow \exists y \ \text{capital}\_of(y,x) ) \land \forall z \ ( \text{capital}\_of(z,x) \rightarrow y = z ) \]  

\text{disjoint( river, mountain )}

\text{capital}\_of \leq_r \text{located}\_in

\text{flow}\_\text{through( domain:river, range:geopolitical\_entity )}

\text{capital} \leq_c \text{city}, \text{city} \leq_c \text{geopolitical\_entity}

\text{c := country := } \langle \text{i(c), } ||c||, \text{Ref}_c(c) \rangle

\{\text{country, nation}\}

\text{river, country, nation, city, capital ...}

OK for learning hierarchy of concepts & relations from text but integration is still hard.
Estevam Hruschka: NELL Never-Ending Language Learner
Humans learn to learn (feedback) – why not machines?

1. Old ML (1.0) had lots of limitations

2. E.g. Without IT memory learned was not cumulative & doesn't leverage past knowledge (human learning is seeding forward)

3. Due to the lack of prior/seed knowledge, ML needs a large number of training examples.

4. Lacks self-learning (w/o supervision) – this needs seeding

5. With ML 1 it is impossible to build a *truly* intelligent system
   - NELL with ML 2 “reads” & has acquired 80 million confidence-weighted beliefs (e.g., servedWith (tea, biscuits)),
Limits of Supervised Learning and ML2?

- Cannot imagine that for every task a large number of training examples need to be labeled by humans.
- ML 2 tries to overcome some of this in a staged curricular fashion, where previously learned knowledge enables learning further types of knowledge.
- *NELL has some degree of self-reflection and the ability to formulate new representations* and
- new learning tasks enable the learner to *avoid stagnation and performance plateaus*...... but while this is a step....
- People still ask,” Is Deep Learning” More Shallow Than We Think? “
  - And if so, what other key capabilities still missing?
Summit Track Themes Are Highly Related

Virtuous Circle provide opportunity: There is synergy among the 3 track areas

Graph Identification++

Seed Knowledge?

dis-organized data

Extractive process

More Intelligent Reasoning
Need comprehensive, integrated meta-knowledge

Reasoning

Need small, ontological building blocks and reference Ontologies?

Opportunity!

From Turning Data into Knowledge using Statistics and Semantics Prof. Lise Getoor
Example: supervised ML framework for predicting phenotype from genotype

M. Yu

Summary

1. The structure of hierarchical systems like the cell can be accurately inferred from data
   1. Enabled by a deluge of experimental measurements
   2. Complements manually curated knowledge (e.g. GO)
2. This structure can be “functionalized” to make biological predictions
   1. Accurate translation of genotype to phenotype
   2. Machine learning guided by ontologies
Some Challenges

How do we:

- bridge between sub-symbolic and symbolic approaches?
- integrate the large range of involved systems & tools?
- create enough quality ontologies to sustain the virtuous circle?
- make learning more intelligent?
- leverage existing ontologies as starter sets for ML?

After Lucas Bechberger “A Bridge Between Neural and Symbolic Representations?”
Example of Intelligent System Challenges:
What still seems missing in ML & neural network AI

- Explainable AI
  - May have to scaffold meta-capabilities like goal, belief, motivation and task driven controlled cognition along with NLU

- A cognitive architecture for intelligence to “emerge from” ingredients:
  - A semi-fixed scaffolding architecture to “develop with”:
    - Some complex combination of bottom-up stimulus driven & top-down goal and knowledge driven processes.
Explainable Artificial Intelligence (XAI)
David Gunning

Black-box:
The model is so complex & the number/space of parameters are so large, it is too difficult to decipher any mechanisms.

Can we make gray or white -boxes? ...or do we design an architecture that does this for us? Explanations lead us into more cognitive systems...

http://www.darpa.mil/program/explainable-artificial-intelligence
What Type of “Better” AI Architectural Boxes Might We Achieve? (Derek Doran, discussed also in Oltimari briefing)

- **Grey-box**: You have some vision/reading/.. mechanisms, but parameters are numerous, decisions are probabilistic, or inputs get "lost" (transformed)
  - Regression, DTs, association rule mining, linear SVMs, clustering....
  - A straight 2D line or rule “explains” some associations but...
- **White-box**: You can see model mechanisms represented simply enough to trace how inputs map to outputs.
- **Note, all these issues about “learning” mechanisms have their analog in the knowledge part of AI architecture.**
  - Humans don’t understand our knowledge or each other fully.
Cognitive Architect as Scaffolding: Popular in Cognitive Psych

Intelligence merges from

- dynamic interaction of goal-directed, task-driven cognition-perception-action cycle
- embodied in our familiar socio-physical environment.

Recap: Mix of Challenge & Opportunities

- Challenges of communicating across the AI, ML, Big Data, Semantic Web, & Applied Ontology disciplines & projects.
  - There remain misunderstandings about what can be accomplished, what the limitations are and how to work.

ML may be a driver & some useful tools have been built.
  - We can leverage some useful best practices (e.g. rich labeling of input features.)

- But there seem practical and foundational challenges (i.e. semantic alignment, handling data & systems heterogeneity & development of reusable building blocks) to make automated K-Acq approaches successful, scalable & robust across/within domains.
Future Prospects
Research with KE associate-like (hybrid) tools to help populate KBs and coordinated ontologies/ontology module from text, data and structured information including LoD and ontologies

New ML/Hybrid systems with capabilities to:

- Interpret/explain their rationale via models,
- characterize their strengths and weaknesses, and
- convey an understanding of how they will behave in the future (DARPA XAI).

Need to investigate the performance-versus-explainability tradeoff space.
But there are Practical Questions too

- How to converge efforts?

- How do we verify and validate the quality of extracted knowledge structures (ontology efficacy)?
  - i.e. if an extraction is created to know about entity, x; who verifies it actually knows, x without inconsistencies?

- If knowledge is extracted from published literature who owns the knowledge/ontology once it is created? Must it be published?

- Can interested parties hijack or compromise a knowledge area?
  - Say by introducing ontological commitments that change the semantics of specific classes or properties in the original source knowledge.

- How do we keep track of conceptual drift in extracted knowledge?