Synthesis "Using Automation and Machine Learning to Extract Knowledge and Improve Ontologies"

Ontology Summit 2017 Track A
AI, Learning, Reasoning, and Ontologies

Gary Berg-Cross, Ontolog Board Member
March 29, 2017
Outline

Vision/Approaches: Automation to Overcome Knowledge Bottleneck for Quality Ontologies

- Session 1 Speakers & Their Topics
  - Alternate approaches but possible hybrid
  - Slowly realizing some reality to the perception of advances in several areas
    - e.g. NLP, information retrieval, ML, data mining, & knowledge representation

- Synthesis
  - Big data & IoT opportunities (NELL etc.)
  - Evolving methods & tools
  - Remaining major challenges of heterogeneity, formalization & integration

3/29/17
Approach to Synthesis of “Automation and Machine Learning to Extract Knowledge and Improve Ontologies”

The range of work springing from automating K-extraction is broad.

• We have leveraged insight from our 5 speakers and the community discussion of approaches, issues and problems.
• We have attempted to distilled the virtual meeting topics to a useful summary for the face-to-face Symposium.
• Our Synthesis is organized into several parts:
  1. Ontology Learning and Population framed by the ontology layer cake framework and status at each level
  2. Approaches
     1. NLP oriented plus
     2. Machine learning plus
     3. Hybrids
  3. Challenges
Our 5 Speakers & Their Talks

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"
Some AI/ML History

• Early AI was optimistic of super intelligence in 20\textsuperscript{th} century (J Sowa)
  – Art Samuel had checker playing system by 1959

• But there remains a knowledge bottleneck
  – Cyc (1984 on) still cannot learn by reading a textbook.
  – But there is good progress in NLP

• And now we have deeper, multi-layer Neural Nets
  – Used to learn time-varying patterns (images, sounds, games)
  – More may be needed
    • Multi-sources of info and hybrid architectures
Estevam Hruschka: NELL Never-Ending Language Learner
Humans learn to learn – why not machines?

Old ML (1.0) had lots of limitations
What it learned was not cumulative

• No memory so learning doesn't leverage past knowledge

Due to the lack of prior knowledge
• ML needs a large number of training examples.
• Lacks knowledge accumulation and self-learning (w/o supervision)

With ML 1 it is impossible to build a truly intelligent system
• Cannot imagine that for every task a large number of training examples need to be labeled by humans
But now with ML-2 many ML Apps

Machine Learning Applications

<table>
<thead>
<tr>
<th>Problem type</th>
<th>Automotive</th>
<th>Manufacturing</th>
<th>Consumer</th>
<th>Finance</th>
<th>Agriculture</th>
<th>Energy</th>
<th>Health care</th>
<th>Pharmaceuticals</th>
<th>Public/social</th>
<th>Media</th>
<th>Telecom</th>
<th>Transport and logistics</th>
</tr>
</thead>
<tbody>
<tr>
<td>Real-time optimization</td>
<td>High</td>
<td></td>
<td>Low</td>
<td>Low</td>
<td>High</td>
<td>Low</td>
<td>Low</td>
<td>Low</td>
<td>Low</td>
<td>High</td>
<td>Low</td>
<td>High</td>
</tr>
<tr>
<td>Strategic optimization</td>
<td>High</td>
<td></td>
<td>Low</td>
<td>Low</td>
<td>High</td>
<td>Low</td>
<td>Low</td>
<td>Low</td>
<td>Low</td>
<td>High</td>
<td>Low</td>
<td>High</td>
</tr>
<tr>
<td>Predictive analytics</td>
<td>High</td>
<td></td>
<td>Low</td>
<td>Low</td>
<td>High</td>
<td>Low</td>
<td>Low</td>
<td>Low</td>
<td>Low</td>
<td>High</td>
<td>Low</td>
<td>High</td>
</tr>
<tr>
<td>Predictive maintenance</td>
<td>High</td>
<td></td>
<td>Low</td>
<td>Low</td>
<td>High</td>
<td>Low</td>
<td>Low</td>
<td>Low</td>
<td>Low</td>
<td>High</td>
<td>Low</td>
<td>High</td>
</tr>
<tr>
<td>Radical personalization</td>
<td>High</td>
<td></td>
<td>Low</td>
<td>Low</td>
<td>High</td>
<td>Low</td>
<td>Low</td>
<td>Low</td>
<td>Low</td>
<td>High</td>
<td>Low</td>
<td>High</td>
</tr>
<tr>
<td>Discover new trends/anomalies</td>
<td>High</td>
<td></td>
<td>Low</td>
<td>Low</td>
<td>High</td>
<td>Low</td>
<td>Low</td>
<td>Low</td>
<td>Low</td>
<td>High</td>
<td>Low</td>
<td>High</td>
</tr>
<tr>
<td>Forecasting</td>
<td>High</td>
<td></td>
<td>Low</td>
<td>Low</td>
<td>High</td>
<td>Low</td>
<td>Low</td>
<td>Low</td>
<td>Low</td>
<td>High</td>
<td>Low</td>
<td>High</td>
</tr>
<tr>
<td>Process unstructured data</td>
<td>High</td>
<td></td>
<td>Low</td>
<td>Low</td>
<td>High</td>
<td>Low</td>
<td>Low</td>
<td>Low</td>
<td>Low</td>
<td>High</td>
<td>Low</td>
<td>High</td>
</tr>
</tbody>
</table>

Source: McKinsey Report 20
Start with parts of speech and phrasing, learn concepts & relations from text etc.

∀x (country(x) → ∃y capital_of(y,x) ∧ ∀z (capital_of(z,x) → y=z))

disjoint(river, mountain)

capital_of ≤_R located_in

flow_through(domain:river, range:geopolitical_entity)

capital ≤_C city, city ≤_C geopolitical_entity

c := country := <i(c), ||c||, Ref_c(c)>

{country, nation}

erver, country, nation, city, capital ...

Associate terms, construct hierarchies & label relations

assign selected terms to an ontological concept.

Concept Hierarchy

Synonyms

Terms

General Axioms

Axiom Schemata

Relation Hierarchy

Relations
Taxonomy learning from text is a challenging task

Taxonomy learning from text can may have several subtasks:

- term extraction,
- hypernym identification and
- taxonomy construction.

Existing approaches for hypernym identification from text rely on:

- lexico-syntactic patterns,
- co-occurrence information,
- substring inclusion, or
- exploit semantic relations provided in text
- definitions

Evaluation Note: In this task, taxonomies are evaluated through comparison with gold standard relations collected from WordNet and other well known, freely available taxonomies.

KB contains all the knowledge accumulated in the past learning of the N tasks. After learning Task N+1, KB is updated with the learned (intermediate as well the final) results from TN+1.

Transfer learning may use Structural correspondence learning. Different applications may have pivot features to chose from.
ML 2 - a systems approach, multiple tasks with multiple learning/mining algorithms

We don't learn in isolation

Semi-Supervised Bootstrap Learning,

We learn effectively from a few examples with the help of the past knowledge.
I knew that, accept it, disagree because...
FRED: Knowledge Extraction based on Discourse Representation Theory and Linguistic Frames

Valentina Presutti, Francesco Draicchio, & Aldo Gangemi (STLab)

They have implemented a novel approach for robust ontology design from natural language texts by combining:

Discourse Representation Theory (DRT), linguistic frame semantics, and ontology design patterns.

Semantic Technology Laboratory
http://wit.istc.cnr.it/stlab-tools/
FRED: Knowledge Extraction based on Discourse Representation Theory and Linguistic Frames

Valentina Presutti, Francesco Draicchio, & Aldo Gangemi (STLab)

They have implemented a novel approach for robust ontology design from natural language texts by combining:

Discourse Representation Theory (DRT), linguistic frame semantics, and ontology design patterns.

Semantic Technology Laboratory
http://wit.istc.cnr.it/stlab-tools/
Main contributions of FRED

- A novel form of machine reading: from text to knowledge graph
- A method for integrating multiple NLP task outputs into a knowledge graph
  - Formal mappings between different theories (implemented as a set of heuristics)
  - Ontology design patterns for the Semantic Web
  - Reference cognitive semantics for interpretation i.e. Frame Semantics

DRT-based frame detection seems feasible after conducting a comparative evaluation of their approach and existing tools.
**Scenario**: examine the relationship between soil microbial communities & trace gas fluxes across the soil-atmosphere interface. Collect all co-located measurements of these 2 kinds of data. I want to limit my DB to only those that were collected in the same ecosystem or vegetation type within 1 mile of each other, and within 2 years of each other.

Are there ontologies & other resources that let me do this?

---

Soil microbial communities are an important component of ecosystem response to environmental change. Abiotic soil properties are measured by observing soil moisture, temperature, pH, and inorganic N.
Using FRED to Fill a Knowledge Gap

Text:

Soil microbial communities are an important component of ecosystem response to environmental change. Production and consumption processes in soils contribute to the global cycles of many trace gases (CH4, CO, OCS, H2, N2O, and NO) that are relevant for atmospheric chemistry and climate. Soil microbial processes contribute substantially to the budgets of atmospheric trace gases. The flux of trace gases between soil and atmosphere is usually the result of simultaneously operating production and consumption processes in soil:

The relevant processes are not yet proven with absolute certainty, but the following are likely for trace gas consumption: H2 oxidation by abiontic soil enzymes; CO cooxidation by the ammonium monooxygenase of nitrifying bacteria; CH4 oxidation by unknown methanotrophic bacteria that utilize CH4 for growth; OCS hydrolysis by bacteria containing carbonic anhydrase; N2O reduction to N2 by denitrifying bacteria; NO consumption by either reduction to N2O in denitrifiers or oxidation to nitrate in heterotrophic bacteria.
Microbial Community associated with trace gases via consumption
Alessandro Oltramari  "From machines that learn to machines that know: the role of ontologies in machine intelligence"

Semantic Deep Learning is new its darkness may give us some deep trouble.

Can we understand how the algorithms generate results?

- Can we understand how the algorithms generate the Results? “It’d be like explaining Shakespeare to a dog”

But....The algorithmic processes are opaque, not the I/O
Handling the Black Box

Reduce the Black Box problem to a more “tractable” issue: how can we clarify the connection between data as input and data as output?

1. **Quantitative methods**: measure the correlation
   - E.g., Family of metrics that capture the degree of influence of inputs on outputs of the system

2. **Semantic transparency**: ontologies of input data (across layered neural networks) can be ultimately interlinked to ontologies of output data
   - E.g. From visual signals, to pixels, bounded boxes, objects, persons, movements, complex activities, etc.
Visual signals are processed by ML classifiers, output is mapped to ontology patterns of action.

Classification of actions is based on semantic structure (frames and roles) and temporal sequences.
Opportunities: Big Data & IoT

Big data provides a great opportunity for LML

• Abundant information from the Web
• Extensive sharing of concepts across tasks-domains
  – Example: natural language learning tasks on different sources are all related

*KDD-2016 Tutorial, Aug 13, 2016, San Francisco, CA, USA*
Evolving Automation

In 2003 36 approaches for ontology learning from text identified. (e.g. ASIUM, Text-to-Onto, Ontolearn and Ontogain)

When reviewed there was:

• a lack of a detailed methodology that guides the ontology learning process from text;

• no fully automated system for ontology learning and many require the involvement of users in order to extract concepts and relations from annotated corpora; and

• a need for a general approach for evaluating the accuracy of ontology learning and for comparing the results produced by different systems.

Evolving Automation -2

Problem: complex or compound sentences may require finding more than one frame or verb class

In response more recent NLP & Ontology Learning have become more cognitive and contextual and leverage Web resources.

“the richness of Web data in terms of (semi-)structured, collaboratively maintained resources, such as Wikipedia, is increasingly being used to improve higher-layer tasks, such as concept formation and relation discovery. We observed from the literature, the current mushrooming of techniques for finding semantic relations using the categorical structure of Wikipedia.”

But there are other natural problems such as how to handle new, but contradictory information (the Web is pretty noisy). A statistical procedure can add the most statistically significant result into a KB, but it hard to clarify afterwards whether the result was appropriate.
Base Ontologies Provide Prior Knowledge to Seed Automation

Global Machine Learning for Spatial Ontology Population
Parisa Kordjamshidi, Marie-Francine Moens
Evolving Automation -3

ML-1 has evolved to ML -2 which is deeper and includes permanent knowledge, but also knowledge management issues.

Both may be needed for a future vision of smart IoT which requires some improved semantics and cognitive modeling.

Valentina thinks that FRED may be used in the NELL pipeline to experiment with a rule-based approach vs. (or combined with) a learning one, for building knowledge graphs with a never-ending approach.

• FRED can benefit from experimenting NELL’s approach for aligning learned knowledge.
• “At the moment we are putting some effort in this direction, but it would be interesting and certainly worth looking into hybridizing the two approaches.”
Big data challenges of automated deployment of IoT apps in unknown IoT environments

To “Know” IoT needs to extract knowledge from raw sensor data. In the future we need to use heterogeneous domain ontologies to semantically annotate data of IoT entities.

Future IoT will have to be dynamic, to a degree, so for putting information in context, there is a major challenge in extracting, and auto-developing knowledge bases/ontologies from a variety of forms.

In the face of big, but noisy data, machine learning will need to show it can build axiom rich ontologies of quality as well as adding light semantics for such related things such as metadata annotation.

It is an open question how long it might take us to automate the harmonization of rationalized ontologies given data-driven concept heterogeneity reflecting the reality of “idiosyncrasies” of particular datasets and vocabularies.