Track B Summary

Co-champions: Mike Bennett, Andrea Westerinen

Summary for Summit Symposium: 15 May 2017
Track B: Using Background Knowledge to Improve Machine Learning Results

• Track Co-champions:
  • Mike Bennett
  • Andrea Westerinen

• Blog Page
Motivations

• Machine Learning (ML) is based on defining and using mathematical models to perform tasks, predict outcomes, make recommendations, etc.

• Initial models can be specified by a data scientist, and/or constructed through combinations of supervised and unsupervised learning and pattern analysis

• Challenges with this:
  • If no background knowledge is employed, the ML results may not be understandable
  • There is a bewildering array of model choices and combinations

• Background knowledge could improve the quality of ML results by using reasoning techniques to select learning models and prepare the training and examined data (reducing large, noisy data sets to manageable, focused ones)
Objective

• The objective of this Ontology Summit 2017 track is to understand:
  • Challenges in using different kinds of background knowledge in machine learning
  • Role of ontologies, vocabularies and other resources to improve machine learning results
  • Design/construction/content/evolution/... requirements for an ontology to support machine learning
Summit Graphical Overview
Track B Positioning
Track B: Sessions

• In the first Track B session on March 15, there were two presentations on combining ontologies with machine learning and natural language processing technologies in order to improve results.
  • In the first case, ontologies were combined with ML to improve decision support. The benefits included improving the quality of decisions, making decisions more understandable, and adapting the decision making processes in response to changing conditions.
  • Regarding combining ontologies with NLP processing, this was in support of digital forensics and situational awareness. Concept extraction from natural language text was improved by using an ontology to isolate the meanings/semantics of the concepts and provide “artificial intuition” into the text.

• In the second Track B session, we aimed to continue discussing the use of ontologies to improve machine learning understandability and natural language processing. We had presentations on:
  • Meaning-Based Machine Learning (MBML), discussing how to get meaningful output from existing machine learning techniques,
  • The use of FIBO and corporate taxonomies to extract and integrate information from data warehouses, operational stores and natural language communications,
  • Driving knowledge extraction via the use of a semantic model/ontology.
Track B Session 1 Content (15 March)

• Simon Davidson (Psonify)
  • The Investigator's Toolkit – Deriving Immediate, Actionable Insights from Unstructured Data

• Ken Baclawski (College of Computer and Information Science, Northeastern University)
  • Combining Ontologies and Machine Learning to Improve Decision Making

• We learned about insight extraction, OODA Loops and situational awareness
Simon Davidson, Psonify

- **Title**: The Investigator's Toolkit – Deriving Immediate, Actionable Insights from Unstructured Data
  - In this session, Simon Davidson of Psonify demonstrates the use of the Investigator's Toolkit for deriving immediate situational awareness and actionable insights from unstructured datasets for forensic investigation. The session starts with a deep dive into the challenges in natural language processing, leading to the provisioning and use of ontologies for the relevant subject matter, as used in the IntuScan platform. As an example, a financial ontology (e.g. FIBO and extensions to that) could be used for financial regulatory compliance challenges, or a legal ontology for analysis of legal texts. This is followed by a live demonstration of the Investigator's Toolkit. The Toolkit is used in digital forensics, extracting concepts with reference to an ontology that isolates the meanings of the concepts and provides “artificial intuition” into the contexts of the subject text, giving immediately actionable insights into the content.
Combining Ontologies and Machine Learning to Improve Decision Making

Decision making is fundamental for modern systems whether they are controlled by humans, computers or both. Machine Learning (ML) is an increasingly popular technique for decision making. This talk will give an overview of work at Oracle and Northeastern University that combines ontologies with ML and other techniques. Benefits of our combination include improving the quality of decisions, making decisions understandable, and improving the adaptability of decision making processes in response to changing conditions. Ontologies are especially well suited to improving decision making that includes both humans and computers. This is the case not only when humans are directly involved as system operators but also when humans are acting as regulators for autonomous systems. These techniques have been applied or proposed in a variety of domains such as customer support, healthcare, cloud services, aircraft operation, and nuclear power plants.
Session 1 Chat Transcript: Key Points

• Requirements: what sort of Ontology?
  • Rigidity etc.

• Clustering versus Hierarchy
  • Suitability for supervised v unsupervised learning
  • Humain in the Loop
  • Requirements for validation of the ontologies used

• Situation awareness and Context
  • Situation Theory

• Ontological formalisms
  • FOL / OWL DL / Higher orders
  • Situation theory needs higher than DL
  • Kinds of reasoning that are appropriate

• Overall
  • Is there a cycle between Track A (synthesizing ontolgiies from data) and Track B (using ontolies in NLP)?
  • Are they the same kinds of ontology?
Track B Session 2 Content (12 April)

• Dr Courtney Falk, Infinite Machines
  • Machine-based Machine Learning

• Bryan Bell, Expert System
  • Leveraging FIBO with Semantic Analysis to Perform On-Boarding, KYC and CDD

• Tatiana Erekhinskaya, Lymba Corporation
  • Converting Text into FIBO-Aligned Semantic Triples

• We learned about phishing, bears and cheese...
Title: Machine-based Machine Learning

Abstract: Meaning-based machine learning (MBML) is a project to investigate how to get meaningful output from existing machine learning techniques. MBML builds from the Ontological Semantics Technology (OST) where natural language resources map to an ontology. An application of MBML to detecting phishing emails provides some initial experimental results. Finally, future research directions are explored.
Bryan Bell, Expert System

• Title: Leveraging FIBO with Semantic Analysis to Perform On-Boarding, KYC and CDD
  • Abstract: The Financial Industry Business Ontology (FIBO) provides a common ontology and taxonomy for financial instruments, legal entities, and related knowledge. It provides regulatory and compliance value by ensuring that a common language can be used for data harmonization and reporting purposes.
  • This session discusses using the formal structure of FIBO and other corporate taxonomies on unstructured information in data warehouses, operational stores and natural language communications (such as news articles, research reports, customer interactions, emails, and product descriptions), in order to create new value and aid in onboarding new customers, establishing a dependable know-your-customer process and complete on-going customer due diligence processes. Financial Industry Business Ontology (FIBO) provides the common language for bridging interoperability gaps and organizing content in a consistent way.
Tatiana Erekhinskaya, Lymba Corporation

• **Title:** Converting Text into FIBO-Aligned Semantic Triples
  
  • **Abstract:** Ontologies are playing a major role in federating multiple sources of structured data within enterprises. However, the unstructured documents remain mostly untouched or require manual labor to be included into consolidated knowledge management process.

  • At Lymba, we are developing a knowledge extraction tool that automatically identifies instances of concepts/classes and relations between them in the text. The extraction is driven by a semantic model or ontology. For example, using FIBO terminology, the system recognizes time/duration constraints in contracts, money values, and their meaning - transaction value, penalty, fee, etc. and links them to the parties in the contract. The extracted knowledge is represented in a form of semantic triples, which can be persisted in an RDF storage to allow integration with other sources, inference, and querying.

  • One more useful add-on is natural language querying capability, when a query like “Find clauses with time constraints for payor” is automatically converted into semantic triples, and then into SPARQL. This talk provides an overview of Lymba’s knowledge extraction pipeline and knowledge representation framework. Semantic parsing and triple-based representation provide a bridge between semantic technologies and NLP, leveraging inference techniques and existing ontologies. We show how Lymba’s Semantic Calculus framework allows easy customization of the solution to different domains.
Session 2 Chat Transcript: Key Points

• Automated learning more contexts
• Identifying the context of a term
  • Linguistic – what part of speech
  • Semantic – what subject area
• Visualization
• Neologisms: how handled?
• Cheese
  • Street name for heroin
  • Attempts to disguise context – illicit activity
• Bear
  • What about a bear market? (sentiment)
• Different tokenizations in different languages
  • Use of a language ontology
• Can we analyze these chat logs with those tools?
Questions?