Global AI Conference

# Day 1

## Using NLP, ML and DL Algorithms to Extract Meaning from Text

@davidtalby

* The Problem
  + Who needs to be vaccinated?
  + Who fits this clinical trial?
  + Who on this protocol did not have this side effect?
  + Who is at risk for sepsis?
  + Who is getting meds they're allergic to?
* At the beginning, there was search
* Then there was semantic search
  + Dictionary Based Attribute Extraction vs. ML Attribute Extraction
* And then, you need to understand language
  + The grammar of sentences
  + Ex. Positive, speculative, possible, negative, conditional, family history, patient history
* 1. Language gets complex and domain specific
  + Human language is wonderfully nuanced
    - Nothing, double negative, compound
    - Speculative: Patient denies alcohol abuse
    - Compound: Allergies - Penicillin, dust, sneezing
    - In reality, a lot of text isn't valid English
* Diagram: Unstructured Information Management Architecture
* 2. You'll need machine learning early
  + Machine Learned Annotators
    - Sometimes easier to just code annotation's business logic
      * Grammatical Patters / Direct Inferences / Lookups
      * If ... then l.. / Age < 18 ==> Child / RIDT (lab test)
    - Sometimes it's easier to learn it from examples
      * Under-diagnosed conditions / Implied by Context
      * Flu, Depression / Relevant labs normal
* 3. Bootstrap and then expand your vocabulary
* Diagram: Unified Medical Language System
* What about expanding & updating ontologies
  + Two reasons
    - 1. Things change
    - 2. Organization => “Humans are an organized species”
* Summary: How / Why
  + 1. Language gets complex and domain specific / Who is at risk for sepsis?
  + 2. You’ll need machine learning early / Who needs to be vaccinated?
  + 3. Bootstrap & then expand your vocabulary / Who fits this culinary trial?

## Keynote Use Case Track: Conversational AI in Amazon Alexa

@ashwinram

<http://linkedin.com/in/ashwinram>

* “The end of touch and the beginning of voice.”
* Timeline
  + 1956 => Type
  + 2007 => Touch
  + 2015 => Voice
* Opportunity and Challenge
  + ESP (Echo Spatial Perception)
    - Which Alexa are you talking with?
  + Alexa, let’s talk about AI
    - Language is hard
      * Speech recognition
      * Language understanding
      * Intent recognition
    - Conversation is harder
      * Goal inference => Why are you saying this?
      * Context modeling => What do we already know?
        + Alexa has to remember shared context
        + What’s 42? => “The answer to the meaning of life, the universe and everything. The trick is knowing what the question is.”
      * Response generation => What should I say in response?
    - The cocktail part problem
* Behind the scenes of conversational artificial intelligence
  + Tech talks available at Amazon for tech talks
* Under the Hood
  + Audio -> Signal processing -> Wake work detection -> Automatic Speech Recognition (Speech to Text) -> Natural Language Understanding -> Artificial Intelligence -> Text to Speech => Answer
* The Alexa Service
* Alexa Ecosystem
  + 1. Skills Kit => For SW Devs
  + 2. Alexa Voice Services => for HW Devs
  + 3. Alexa Fund => for Entrepreneurs
* Amazon Alexa Prize
  + $2.5M competition to advance the field for conversational AI
  + Challenge
    - Create a socialbot that can converse coherently and engagingly on popular topics for 20 minutes
    - Next November => Finals
    - March => Releasing skills used by competitors

Lunch Break

Use Case Track: AI in Medicine

* "The problem with AI isn't AI it's AB."
* "What is it like at Stanford with all these brilliant minds in Computer Science?"
* Need to think of AI as Deep Learning, Natural Learning
* AI/The Paast
  + Turing
  + Dartmouth Conference
  + The Perceptron
    - Huge gap between AI and doctors/nurses
  + AI's winters: 60s & 70
  + 90s: AI took hold as a legitimate concentration
  + "Rule-Based Expert Systems"
  + \*\* Despite these Expert Systems good at rule-based, the decision-making was not in-sync with the doctors' workflow
  + Computer beat 90% of IT specialists on infectious diseases
* "Concatenation"
* AI/The Present
  + Data -> Information -> Knowledge -> Intelligence
  + Why AI and Healthcare now than 5 years ago?
    - The amount of data available did not exist 5 years ago
    - Moore's Law -> Computational Power
    - "Trifecta in AI & Healthcare"
  + Problems with Data & Healthcare
    - 1. Data very unstructured => Critical for data manipulation
      * Unstructured Data to Results => Unstructured Data -> Structured Data -> Source Data -> Intermediate Data -> Resultant Data
    - 2. Heterogeneity and Volume of Data
    - 3. For Clinicians: Healthcare data is doubling every 18 months
      * No living Dr. That can incorporate that much data from all the published reports w/in reason
    - 4. Coming Data Sources => "Healthcare Tsunami"
    - BIGGEST PROBLEM: Data is very sequestered in individual hospitals
  + The potential solution => BitCoin/blockchain technology in healthcare
  + Most DRs haven't heard of supervised/unsupervised
    - Have them work with Data Scientists who know that & find solutions in healthcare
  + Hot topics
    - Insights
    - Imaging and diagnostics
    - Wearable technology
  + Six Areas of
    - 1. Decision Support & Hospital Monitoring
    - 2. Medical Imaging & Biomedical Diagnostics
    - 3. Precision Medicine & Drug Discovery
    - 4. Cloud Computing & Big Data
    - 5. Digital Medicine & Wearable Tech
    - 6. Robotic Tech & Virtual Assistants
  + In the future, you will find Data Scientists "rounding" decisions in the hospitals
  + Stanford
    - Python in Project
    - "Very important to know the tools"
  + Google used Deep Learning diagnose and analyze medical imaging
    - Verily, Talk to Tim
  + "Convergence" => one word to describe everything
  + "AI makes the visible invisible and the invisible visible." - Dr. Chang
* AI/The Future
  + Healthcare is still talking about RDBMS
  + "The Laws of Medicine", Mukherjee
    - We don't learn enough from the end of one patient
  + "One shot learning" => taking data from one sample size and learning a lot from it
  + Create a medical brain for all the subspecialties in medicine
    - Learn from previous mistakes
    - Decision support
    - AI methodologies => medical brain => FOR EVERYONE!
  + "Convolution"
    - Need data scienctists working with clinically trained data scientists
    - Cannot just have clinicians work in silos
    - Need a convergence/convolution of the two fields
  + "AI is the orchestra, but we need conductors"
  + Educate the clinicians that it's the new energy source to illuminate the dark spots in medicine.
* AI Med

Technical Track: Machine Learn at Scale

* Operator
  + Basic unit of processing
* PRC
  + Productivity => UI, Analysis, Searching, Advanced Scheduling
  + Reusability => Operators, Operator Library, Workflow
  + Comprehensiveness => Type system, Toolset support
* 25% FB engineers / 1.7B People
* research.facebook.com

Day 2

AI Combined with Deep Learning

Bryan Bell, Executive VP, Expert System

* Proven solution combining AI on semantics combine DL. A system that understands text that mimics humans
* Agenda
  + Demo real time content analysis
    - <http://intelligenceapi.com>
  + What is semantics & its importance?
    - Demo machine’s ability to understand word context
    - The Disambiguator
  + Use Case: Customer support demo => Answering 24kk questions annually => $7 savings per question
  + AI combined w/ DL makes smarter machine
    - <http://cogitoapi.com/demo>
  + At the heart of solving a problem is the ability for our brain to understand the info being analyzed
    - Brain needs to understand disambiguities
* Objective of AI: Build a system that understands text in a way that mimics the human ability to comprehend info
  + Goal
    - Look at content
    - Establish meaning
    - Tag it
    - Combine with DL => something powerful!
* Why is word disambiguation required? => Without word context, meaning is lost
  + Many words with multiple meanings
    - Jaguar => cat, car, OS
  + Different words with same meaning
  + Same word with similar meanings
    - Organization => charity, company, non-profit
* Semantic Network => Ontology of knowledge (knowledge graphe)
  + Multi-dim map of mean ingful words w/in a lang, along w/ defs of those words & poss rel’s b/w those words
* Knowledge Graph
  + Enables
    - Index large vols of unstr data
    - Analyze data set
    - Est word context
    - Can be enriched w/ new domain specific knowledge as required
    - No training necessary
* Linguistics analysis combined w/ semantics reasoning

|  |  |  |
| --- | --- | --- |
| Reqs | Def | Ex |
| Morphological Analysis | Understand word forms | Dog, dog-catcher, doggy-bag => closely related |
| Grammatical Analysis | Understand the parts of speech | Noun vs verb usage of words |
| Logical analysis | Understand how words relate to other words |  |
| Semantic analysis (dsambig) | Understand words in context |  |

* Powerful & Intelligent
  + KG must be:
    - Optimized
    - Easily customized
    - Acquires new knowledge
  + **Obj**: Combine AI w/ DL
* RT Content Analysis
* Customer Interaction => Connect customers with the right information

## Use Case Track: Using NLP to Diagnose and Address Logistics Challenges in Global Pharmaceutical Supply Chains

* Thematix Partners => <http://thematix.com>
  + Business architecture
  + Ontology engineering
  + Semantic Technologies – architecture, design training
  + Applied semantics for knowledge and SEO
* Diagram: The Unstructured Info Availability Problem
* Problem: Bridge the holes between unstructured and structured data
* EX: EpiPen
* NLP
  + Relevance Analysis
  + Sentiment Analysis
  + Text Extraction
  + Document Summarization
  + Named Entity Recognition
  + Document Retrieval
  + Document Translation
  + Technology has been around for a long time
    - Did not have compute power
    - Challenges between time & Machine learning
    - What’s changed
      * Ability & ability to leverage machine learning
      * Methods to understand and break down to understand what it says has been around since the 80s
* OpenNLP => Not for the faint-hearted but it’s free!
* Semantic Analysis
  + Reqs knowledge about words/phrases and defs in context
  + Requires general knowledge about the world (exp’d in semantic lang)
  + Reqs spec knowledge about the world (exp’d in sem lang)
  + Knowledge about word order and syntax
  + Reference rules => What entity does \_\_ refer to?
  + Rules of discourse
  + Ability to check if semantic interpretation is consistent w/ the general and spec knowledge about the world
  + Corner case: “Subject to object raising”
    - Ex: The unicorn is behind the tree
    - To move at a level of abstraction => The unicorn appears to be bhind tree
    - Higher abstraction => It seem there is a unicorn behind the tree
    - Keep raising abstraction & see if the system can figure it out
* Ontologies, Knowledge Graphs & the Semantic Web
* What is ontology?
  + Rich description of the …
    - Terminology, concepts, nomenclature
    - Relationships among concepts and indivs
    - Sentences distinguishing concepts, refining definitions and relationships (constraints, restrictions, regular expressions)
  + … relevant to a particular domain or area of interest
* Ontology-based technologies
  + Ont provides common vocab for use by resources, processes, services
* Knowledge graphs, databases and ontology
  + Graph databases!!
* Supply Chain Issue Analysis
* OpenNLP Language Models
  + Sentence models
  + Token models
  + Named entity recognition
  + Part of Speech (PoS) tagging
  + Chunking
  + Parsing
* Diagram: Statistical NLP Architecture
* Example Issue Description
* Diagram: Semantic Transformation & categorization
* Diagram: Applying VerbNet Syntactic Templates
* Diagram: VerbNet Semantic Links to Ontology
* Diagram: Convert Parse to Assertions for Validation
* SPARQL Queries for Analysis => Select … Where { … }

NASA Defend the World with AI

* DEFLECTOR SELECTOR. All you need to know!!

AI Based Data Extraction

* Data Assets Classification
  + Semi-structured
  + Unstructured
    - Classifiable + Known
    - Classifiable + Unknown
  + Structured
* The Dark Side of Unstructured Data
* The Unstructured Data Prep Challenge
  + Unstructured Data -> Data Extraction -> Data Transformation -> Bring data to be processable
* Classifiable w/ Known Content
* Extractors
  + Learning Extractors => human in the loop machine learning (proprietary algorithm) + atomic and composed fields
  + Value Matching Extractors => find a set of values in a certain context
  + Visual Pattern Extractor => supervised ML
  + Table Extractors => supervised ML
  + Semantic Extractors => Boolean and list + need NLP scripting
  + Packaged Form Extractors => IRS Forms, etc
* The Extractor Marketplace
  + AI Based Extra
    - Learning Extractor
    - Visual Pattern Extractor
    - Table Ex
    - Semantic Boolean
    - Semantic List
  + REGEX
    - Value Matching
    - List
    - Other
  + FORM
    - SEC Form ?.. etc etc
* Field Types with their Extraction Methods
* ETI – Extract Transform Integrate
* Platform – Human in the loop ML
* Controlled Natural Language (CNL)
* TABSTRACT
* Diagram: The Smart Data Extraction and Integration Platform
* Recognos => <http://www.recognos.com/>

Fintech AI

@natbusa

* Predictive APIs: How to get tthere
* Rule-based Sys
* Classic Machine Learning
* Rep ML
* DL (e-e learning)
* From Feature to Architecture Eng
  + Traditional ML Flow: I -> Feat Ext -> Features -> Trad ML Algo -> O
  + DL Flow: I -> DL Algo -> O
* Step 3:
* Step 4: Picking the brain of our DL Model => Each time you compute thru a layer, you get an embedding
* Step 5: Semantic Clustering => tagging
  + Reverse eng the NN and make a formula
  + Extra level of transparency can be very beneficial
* Hands on Kera and Tensorflow
* Hyper-Parameters Tuning
  + <https://arxiv.org/abs/1611.03824v1>
  + "Efficient and Robust Automated Machine Learning", Feurer
    - <https://papers.nips.cc/paper/5872-efficient-and-robust-automated-machine-learning>
* Network Intrusion Detection
  + Techniques: TDA, Dimensionality Reduction
* Main Point #1: AI is about Tools and Technology
  + Spark
  + TensorFlow
  + Sci-kit Learn
  + Jupyter
* Main Point #2: AI, An ensemble of analtical methods
  + SQL + Graph + Text + Machien Learning + Voice/Image/Video
* Recap
  + 1. Use all the data
  + 2. AI Can be applied in Finance: Yes
  + 3. AI: Tools, Use Cases, Data, Experts

Deep Learning Frameworks

* Introduction
  + NN OS'd
  + Topics Discussed
    - NN
    - Uses of NN
    - TensorFlow
    - Torch
    - CNTK
    - Caffe
    - Theano
    - Comparative metrics
    - Further Reading
* DL Architectures
  + Deep NN
  + Deep Belief Networks (DBNs)
  + Recurrent NN (RNNs)
  + Convolutional NN (CNNs)