Natural Language Processing

State of the Art and Possible Directions

Nari Kannan

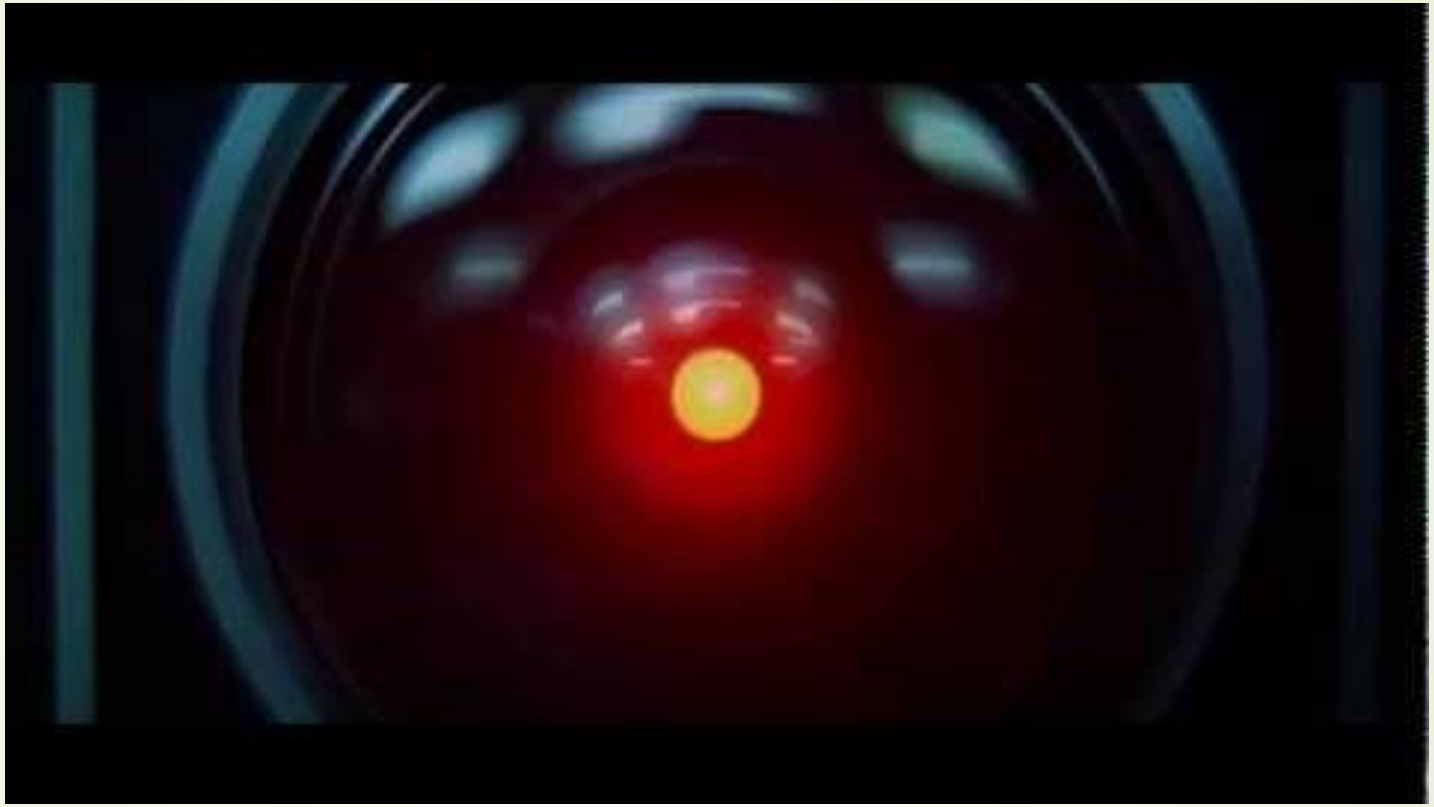
Bethesda Artificial Intelligence Meetup

8/3/2019

Introduction

- **Nari Kannan**
- Over 3 Decades of Programming, Software Development, Management
- 1983-1985 – Natural Language Processing Research Group – University of Massachusetts at Amherst
- Master's Project – **A Natural Language Interface to the VAX/VMS Operating System**
- 10 Years at Digital Equipment Corporation's **AI Applications Group**
- **Patent** for method that used statistics and word vectors to send customer support emails to the right group

Scene from 2001:A Space Odyssey



Click Here if the YouTube Link does not work above – and Remember to Turn CC On to see how Captioning does not work well

Dialog from 2001:A Space Odyssey

DAVE: Open the pod bay doors, Hal.
HAL: I'm sorry, Dave. I'm afraid I can't do that.
DAVE: What's the problem?
HAL: I think you know what the problem is just as well as I do.
DAVE: What are you talking about, Hal?
HAL: This mission is too important for me to allow you to jeopardize it.
DAVE: I don't know what you're talking about, Hal.
HAL: I know that you and Frank were planning to disconnect me, and I'm afraid that's something I can't allow to happen.
DAVE: Where the hell'd you get that idea, Hal?
HAL: Although you took very thorough precautions in the pod against my hearing you, I could see your lips move.
DAVE: All right, Hal. I'll go in through the emergency air lock.
HAL: Without your space helmet, Dave, you're going to find that rather difficult.
DAVE: Hal, I won't argue with you anymore. Open the doors!
HAL: Dave...This conversation can serve no purpose anymore. Goodbye.

Dialog from 2001:A Space Odyssey

Anaphora – "that" is "open"

Inductive Reasoning

Self-Awareness Self-Preservation

Deductive Reasoning

DAVE: Open the pod bay doors, Hal.
HAL: I'm sorry, Dave. I'm afraid I can't do that.
DAVE: What's the problem?
HAL: I think you know what the problem is just as well as I do.
DAVE: What are you talking about, Hal?
HAL: This mission is too important for me to allow you to jeopardize it.
DAVE: I don't know what you're talking about, Hal.
HAL: I know that you and Frank were planning to disconnect me, and I'm afraid that's something I can't allow to happen.
DAVE: Where the hell'd you get that idea, Hal?
HAL: Although you took very thorough precautions in the pod against my hearing you, I could see your lips move.
DAVE: All right, Hal. I'll go in through the emergency air lock.
HAL: Without your space helmet, Dave, you're going to find that rather difficult.
DAVE: Hal, I won't argue with you anymore. Open the doors!
HAL: Dave...This conversation can serve no purpose anymore. Goodbye.

Agenda

- Natural Language Understanding vs Natural Language Processing Vs Speech Processing
- Syntax Vs Semantics
- Syntactic Approaches
- Semantic Approaches
- Scripts, Plans, Goals
- Tricky Natural Language Processing Issues
- Tricky Voice Recognition Issues
- Main Takeaways

Natural Language Processing (NLP).VS. Natural Language Understanding (NLU)

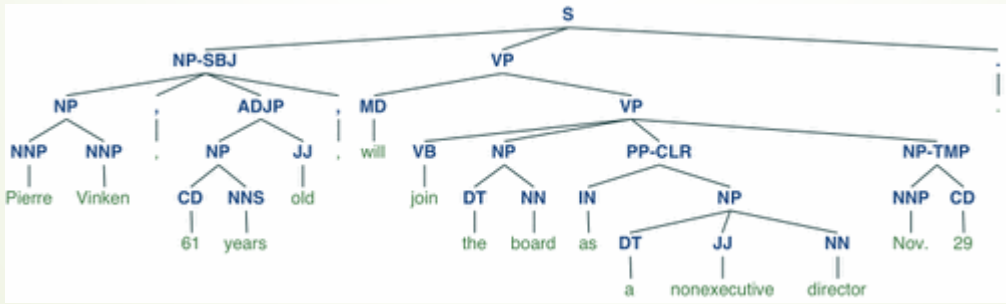
- **Natural Language Processing (NLP)** - Systems that work together to handle end-to-end interactions between machines and humans in the preferred language of the human. In other words, NLP lets people and machines talk to each other more naturally.
  - Narrow, Goal-Oriented, Not Too Intelligent, Can Still Accomplish A Lot!
- **Natural Language Understanding (NLU)** addresses how to best handle unstructured inputs that are governed by poorly defined and flexible rules and convert them into a structured form that a machine can understand and act upon.
  - Deeper understanding of what is being said, involves higher levels of reasoning, "Broader Intelligence"
- **Speech Processing** has steps **BEFORE** NLP/NLU and then **AFTER**
  - **Speech Recognition->NLP->(NLU) ->Execute Actions (Search, Physical Actions like Switching Lights ON/OFF)->Verbal Response->Speech Synthesis**

Syntax Vs Semantics

- **Syntax** is the study of the **structure of sentence**
- **Semantics** is the study of **meaning**
- **Good Approaches need to combine both to increase performance**

Syntactic Approaches

- Using Toolkits for Syntactic Analysis e.g. NLTK
  - **Tokenization** - ['At', 'eight', "o'clock", 'on', 'Thursday', 'morning', 'Arthur', 'did', "n't", 'feel', 'very', 'good', '.']
  - **Identify Grammatical Structures** - Tree('S', [('At', 'IN'), ('eight', 'CD'), ("o'clock", 'JJ'), ('on', 'IN'), ('Thursday', 'NNP'), ('morning', 'NN'), Tree('PERSON', [('Arthur', 'NNP')]), ('did', 'VBD'), ("n't", 'RB'), ('feel', 'VB'), ('very', 'RB'), ('good', 'JJ'), ('.', '.')])
  - **Prepare Parse Tree**



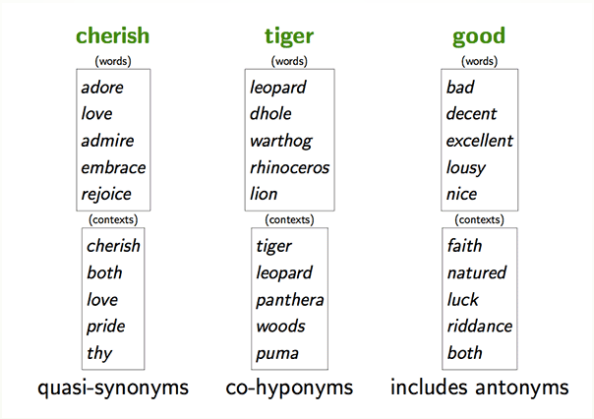
  - **Some other things Syntactic Toolkits Can Do**
    - **Sentiment Analysis –** What's the tone? – Angry, Sad, Indifferent, etc.

Other Syntactic Toolkits

- Stanford's CORE NLP Toolkit
- Sharp NLP – C# Based Toolkit
- Other Toolkits – Information Retrieval, Machine Translation Toolkits

Semantic Approaches to NLP

- Many ways to classify Semantic Approaches. Here's one:
  - **Distributional Approaches**
    - **Use of Large-scale Statistical Machine Learning and Deep Learning** - word vectors for mathematical analysis - Which words went with which others in various contexts – Use of Corpora – Social, Novels, News Reports, Essays, SMS and Text Messages, etc.



  - **Frame Based**
    - **Expectation Driven –** Context Frames are instantiated and confirmed – More on this in Scripts, Plans, Goals discussion

Semantic Approaches to NLP

- **Model-Theory Based**
  - Use Higher-Order Logic used to model what is being said or asked

- **Interactive Learning Based**
  - Language as a cooperative interactive learning game between speaker and listener

Semantic Toolkits

- General Architecture for Text Engineering (GATE)
- RapidMiner Text Mining Extension
- KH Coder
- Google Cloud Natural Language API
- VisualText

Scripts, Plans and Goals Research

- **Yale** School of Natural Language Processing Research
- **Tenets of this approach** – Semantic Analysis and Deep Reasoning is needed for ***"filling in things that are not explicitly mentioned"***
- **Expectation-driven Frame-Based Parsing** – When you encounter certain words or phrases, you instantiate all frames that fit and keep those that you additional evidence of!
- ***"The waiter took a long time bringing the menus"***
  - Assumes a **Restaurant** Script – You enter a Restaurant. You get Seated. They Bring you Menus. You Order. They bring you food. You Eat. The Waiter brings the check. You Pay. You leave.
  - When parsing the above sentence, instantiate the Restaurant Frame when Menus are mentioned. "Waiter" re-inforces it. It is not a computer pull down menu that is meant
- **Plans** and **Goals** further refine this approach with Plan Frames and Goal Frames that are attached to Scripts

Tricky Natural Language Processing Issues

- **Ambiguity** – "I heard his cell phone ring in my office" – Is his cellphone in my office or is he remote and calling my office from his cell phone?
- **Coherence** – "John likes Bill. He gave him an expensive Christmas present" – who got the present – John or Bill?
- **Co-Reference** – "**Our neighbors** dislike the music. If **they** are angry, the cops will show up soon." – They referring to Our Neighbors need to be resolved
- **Personality, Intention and Style** – "Oh. Great!". – Is this being happy or sarcastic?
- **Idioms –** Idioms are tricky to handle – "He hit one out of the ballpark" – No real baseball may be involved with this one!
  - **French – "I have other cats to whip"** meaning **"I have other things to do"**
  - **Portuguese – "Take your little horse away from the rain"** meaning **"Don't hold your breath – it's not going to happen"**

Tricky Voice Processing Issues

- **Background Voices**
- **Accents**
- **Speaker's distance from the Phone**
- **Pauses**…"Um"s and "Ah…"s
- **Similar Sounding Words** – "Let's meet up Tuesday" interpreted as "Let's meet up today"

Main Takeaways

- Natural Language Processing works well for simple, syntax-based approaches
- Natural Language Understanding gets into meaning of things, deeper reasoning, and use of logic
- Many toolkits available for task-oriented interfaces – Commands, Home IoT, Search and simple Question-Answering
- For better Conversations, Semantic approaches may be necessary
- Toolkits are also be available for Semantic Analysis and inclusion for use along with Syntactic Toolkits

If you want to dig deeper….

- Speech and Language Processing – Martin Jurafsky – Free PDF, 975 pages
- Natural Language Processing: A Quick Introduction to NLP with Python and NLTK (Step-by-Step Tutorial for Beginners) -Samuel Burns
- Natural Language Processing with TensorFlow: Teach language to machines using Python's deep learning library - Thushan Ganegedara
- Natural Language Understanding (2nd Edition) - James Allen
- Scripts, Plans, Goals, And Understanding: An Inquiry Into Human Knowledge Structures – Roger Schank, Robert P. Abelson

Questions?