RIPS 2018 Readings

Table of Contents

General Machine Learning and Artificial Intelligence

Andrej Karpathy Blog

- Hacker's guide to Neural Networks
- The Unreasonable Effectiveness of Recurrent Neural Networks

i am trask

I recommend starting here for basic Neural Network concepts.

- A Neural Network in 11 lines of Python (Part 1)
- A Neural Network in 13 lines of Python (Part 2 - Gradient Descent)
- Anyone Can Learn To Code an LSTM-RNN in Python (Part 1: RNN)

WildML

- Recurrent Neural Networks Tutorial, Part 1 – Introduction to RNNs
- Recurrent Neural Networks Tutorial, Part 2 – Implementing a RNN with Python, Numpy and

Theano
- Recurrent Neural Networks Tutorial, Part 3 – Backpropagation Through Time and Vanishing Gradients
- Recurrent Neural Networks Tutorial, Part 4 – Implementing a GRU/LSTM RNN with Python and Theano

Other Sources

- Awesome Machine Learning
- Understanding LSTM Networks

Knowledge Graphs and Computational Fact-Checking

Knowledge Graphs are an increasingly popular data structure for representing relational information. They assume a knowledge base in the form of relational triples (subject, predicate, object) and model these triples using a graph an ordered pair G=(V,E) where V is a set of concept nodes and E is a set of predicate edges.

- Automated Fact-Checking presentation by Joshua Chen

Papers

- Computational Fact Checking from Knowledge Networks
- Computational Fact Checking through Query Perturbations
- Discriminative Predicate Path Mining for Fact Checking in Knowledge Graphs
- Open Information Extraction: The Second Generation
- A Review of Relational Machine Learning for Knowledge Graphs
- Towards Computational Fact-Checking

Natural Language Processing

Natural language processing (NLP) is an area of computer science and artificial intelligence concerned with the interactions between computers and human (natural) languages, in particular how to program computers to process and analyze large amounts of natural language data. Challenges in natural language processing frequently involve speech recognition, natural language understanding, and natural language generation. I recommend starting with "Text Mining: the State of the Art and the Challenges" for an overview of text mining.

General

Articles, Blogposts, and Tutorials

- edX Course on Natural Language Processing
- Oxford Deep NLP 2017 Course
- Regular Expressions 101

Papers

- Evolving Better Stoplists for Document Clustering and Web Intelligence
- On Stopwords, Filtering and Data Sparsity for Sentiment Analysis of Twitter
- Preprocessing Techniques for Text Mining - An Overview
- Retrieval Effectiveness on the Web
- Risk Information Extraction and Aggregation
- Text Mining: The State of the Art and the Challenges

word2vec and doc2vec

Efficient Estimation of Word Representations in Vector Space and *Distributed Representations of Words and Phrases and their Compositionality* started it all. Here are the links for documentation on word2vec and doc2vec.

Articles, Blogposts, and Tutorials

- A Gentle Introduction to Doc2Vec
- Vector Representations of Words
- Word2Vec Tutorial - The Skip-Gram Model
- Word2Vec Tutorial Part 2 - Negative Sampling

Papers

- An Empirical Evaluation of doc2vec with Practical Insights into Document Embedding Generation
- Distributed Representations of Sentences and Documents
- Distributed Representations of Words and Phrases and their Compositionality
- Efficient Estimation of Word Representations in Vector Space
- Neural Network Doc2vec in Automated Sentiment Analysis for Short Informal Texts

Parsing to Logical Form

The problem of learning to parse sentences to lambda-calculus representations or Subject-Predicate-Object triples of their underlying semantics.

First-Order Logic

Papers

- [Online Learning of Relaxed CCG Grammars for Parsing to Logical Form](#)

Subject-Predicate-Object Triples

Articles, Blogposts, and Tutorials

- [Reverb](#)

Papers

- [From Information to Knowledge Harvesting Entities and Relationships from Web Sources](#)

Relational Learning

Statistical relational learning (SRL) is a subdiscipline of artificial intelligence and machine learning that is concerned with domain models that exhibit both uncertainty (which can be dealt with using statistical methods) and complex, relational structure. Note that SRL is sometimes called Relational Machine Learning (RML) in the literature. Typically, the knowledge representation formalisms developed in SRL use (a subset of) first-order logic to describe relational properties of a domain in a general manner (universal quantification) and draw upon probabilistic graphical models (such as Bayesian networks or Markov networks) to model the uncertainty; some also build upon the methods of inductive logic programming.

General

Papers

- [Community Detection in Graphs](#)
- [Philosophers are Mortal: Inferring the Truth of Unseen Facts](#)
- [A Review of Relational Machine Learning for Knowledge Graphs](#)

Managing Relational Data

Resource Description Framework

- [Getting Started with RDFLib](#)
- [RDF 1.1 Concepts and Abstract Syntax](#)
- [Resource Description Framework](#)

SPARQL Protocol and RDF Query Language

- Querying with SPARQL
- SPARQL Endpoint Interface to Python
- SPARQL Protocol and RDF Query Language
- SPARQL Wrapper