

RIPS 2018 Readings

Table of Contents

- General Machine Learning and Artificial Intelligence
- Natural Language Processing
 - General
 - Information Extraction
 - word2vec and doc2vec
- Semantic Web
 - General
 - Knowledge Graphs
 - Computational Fact-Checking
- Statistical Relational Learning
 - General
 - Managing Relational Data
 - Resource Description Framework
 - SPARQL Protocol and RDF Query Language

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](#)

Natural Language Processing and Information Extraction

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

- [Accurate Unlexicalized Parsing](#)
- [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](#)

Information Extraction

Information extraction (IE) is the task of automatically extracting structured information from unstructured and/or semi-structured machine-readable documents.

Articles, Blogposts, and Tutorials

- [Reverb](#)

Papers

- [From Information to Knowledge Harvesting Entities and Relationships from Web Sources](#)
- [Identifying Relations for Open Information Extraction](#)
- [OpenIE-based approach for Knowledge Graph construction from text](#)
- [Open Information Extraction from the Web](#)
- [Open Information Extraction: The Second Generation](#)
- [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](#)

Semantic Web

The Semantic Web can be thought of as a “web of data.” The ultimate goal of the Semantic Web is to enable computers to do more useful work and to develop systems that can support trusted interactions over the network. The term “Semantic Web” refers to W3C’s vision of the Web of

linked data. The trend toward the Semantic Web is based on the goal of semantic interoperability of data, which enables application independence, improved search facilities, and improved machine inference.

General

Articles, Blogposts, and Tutorials

- [DaConta's *Semantic Web*](#)
- [Wikipedia's Page on Semantic Web](#)

Knowledge Graphs

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.

Articles, Blogposts, and Tutorials

- [Automated Fact-Checking presentation by Joshua Chen](#)

Papers

- [Computational Fact Checking from Knowledge Networks](#)
- [Discriminative Predicate Path Mining for Fact Checking in Knowledge Graphs](#)
- [OpenIE-based approach for Knowledge Graph construction from text](#)
- [A Review of Relational Machine Learning for Knowledge Graphs](#)
- [Towards Computational Fact-Checking](#)

Computational Fact-Checking

Articles, Blogposts, and Tutorials

- [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](#)
- [Towards Computational Fact-Checking](#)

Statistical 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

The Resource Description Framework is an XML-based language to describe resources. Please note: RDFs use URIs (Uniform Resource Identifiers) so rather than seeing something like "tree bark" in an RDF you will see something like "www.example.org/ontology/plant/tree/#bark." Additionally, there are two kinds of notations for RDFs, the subject-predicate-object notation, called **N3** notation, which is more intuitive and the **serialized** format which is all of the triples aggregated and nested in a huge mess. To give an example, consider the following three sentences:

- Buddy Belden owns a business.
- The business has a Web site accessible at <http://www.c2i2.com/~budstv>.
- Buddy is the father of Lynne.

The N3 or subject-predicate-object notation is:

- <#buddy> <#owns> <#business>.
- <#business> <#has-website> <http://www.c2i2.com/~budstv>.
- <#buddy> <#father-of> <#lynne>

The serialized RDF looks like this:

```

<rdf:RDF
  xmlns:RDFNsId1='#'
  xmlns:rdf='http://www.w3.org/1999/02/22-rdf-syntax-ns#'>
  <rdf:Description rdf:about='#Buddy'>
    <RDFNsId1:owns>
      <rdf:Description rdf:about='#business'>
        <RDFNsId1:has-website
rdf:resource='http://www.c2i2.com/~budstv' />
      </rdf:Description>
    </RDFNsId1:owns>
    <RDFNsId1:father-of rdf:resource='#Lynne' />
  </rdf:Description>
</rdf:RDF>

```

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

Data Sets:

- [Freebase](#)
- [Linked Open Data Cloud](#)
- [PubChem](#)

SPARQL Protocol and RDF Query Language

- [Querying with SPARQL](#)
- [SPARQL Endpoint Interface to Python](#)
- [SPARQL Protocol and RDF Query Language](#)
- [SPARQL Wrapper](#)