(Brief) Introduction to (Selected Aspects of) Natural Language Processing and Machine Learning

Jaromir Savelka and Matthias Grabmair

Intelligent Systems Program
University of Pittsburgh

jas438@pitt.edu

ASAIL 2015, San Diego, CA
June 12, 2015

Presentation Overview

- Intuition
  - Natural Language Processing
  - Machine Learning
  - Natural Language Processing and Machine Learning
- Natural Language Processing for Machine Learning
- Machine Learning for Natural Language Processing

Presentation Overview

Natural Language Processing (NLP)

Natural language processing is field of study focused on what goes into getting computers to perform useful and interesting tasks involving human languages.

NLP is also concerned with insights that such computational work gives us into human processing of language.

There is strong AI aspect to NLP:

- ill-defined problems,
- non-feasibility of exact solutions, etc.

... demanding specific paradigms for approaching the tasks, e.g.:

- state-space search (to manage problem of making choices)
- dynamic programming (to avoid having to redo work)
- machine learning classifiers (decisions based on features extracted from local context)

[6] Litman 2013

Natural Language Processing (cont.)

Major Topics

- words
- syntax
- semantics
- pragmatics

Example Big Applications

- Question Answering
- Conversational Agents
- Summarization
- Machine Translation

[6] Litman 2013

Presentation Overview

Machine Learning (ML)

The field of machine learning studies the design of computer programs (agents) capable of:

- learning from past experience or/and
- adapting to changes in the environment.

Learner (ML system) processes data $\boldsymbol{D}$ representing past experience and tries to either:

- develop appropriate response to future (not known at time of learning) data, or
- describe $\boldsymbol{D}$ in some meaningful way.

[6] Hauskrecht 2015

Machine Learning (cont.)

- Supervised learning
  $\boldsymbol{D} = \{d_1, d_2, \ldots, d_n\}$ where $d_i = \langle \boldsymbol{x}_i, y_i \rangle$
  $\boldsymbol{x}_i$ is numeric input vector (features);
  y_i is desired output (label)
  objective is to learn $f : \boldsymbol{x}_i \rightarrow y_i$
- Unsupervised learning
  no y_i
  objective is to learn relations among individual $\boldsymbol{x}_i$
- Reinforcement learning

Three Basic Steps in Learning

1. select function $f(\boldsymbol{x}_i)$ (model)
2. select error function
3. find parameters for $f(\boldsymbol{x}_i)$ optimizing error function

[6] Hauskrecht 2015

Presentation Overview

Natural Language Processing and Machine Learning

ML and NLP are closely related, often overlapping, fields of study (system can often be understood as both, NLP and ML).

Main distinction is in goals NLP and ML want to achieve.

Typical interactions of NLP and ML:

- use of NLP techniques to transform natural language text into $\boldsymbol{D}$ for training of ML system
- use of ML techniques to help NLP system achieve its goal

Basic Setup for NLP-ML Experiment

1. Take $\boldsymbol{D}$ (annotated? corpus) and divide it into:
   - training data set $\boldsymbol{D}_{train}$, and
   - testing data set $\boldsymbol{D}_{test}$.
2. Select a model or set of models.
3. Select error function.
4. Find parameters for the model optimizing the error function.
5. Apply the learned model to $\boldsymbol{x}_i \in \boldsymbol{D}_{test}$.
6. Compare predicted labels with $\boldsymbol{y}_{test}$.
   Results of evaluation can be used to compare different models to determine the best one.

Presentation Overview

Intuition
- Natural Language Processing
- Machine Learning
- Natural Language Processing and Machine Learning

Natural Language Processing for Machine Learning

Machine Learning for Natural Language Processing

Transformation of Corpus into Vector Space Model (VSM)

- Our raw data is large corpus of natural language text.
- We transform the raw data using desired linguistic processing:
  - tokenization (extraction of terms from raw text)
  - normalization (drop superficial variations)
  - annotation (introduce additional variations)
- Build frequency matrix (document-term matrix).
- Weight values in the matrix?
- Smooth the matrix?

[4] Turney & Pantel 2010

Transformation of Corpus into VSM (cont.)

Raw data: This ruled those rules out of the consideration.

Tokenization: {ruled, rules, out, consideration}

Normalization: {rule, rule, out, consideration}

Annotation: {rule/VBN, rule/NNS, out/IN, consideration/NN}

Transformation of Corpus into VSM (cont.)

Frequency Matrix

...	rule/VBN	...	rule/NNS	...	out/IN	...	consideration/NN	...
...	1	...	1	...	1	...	1	...

Weighting (TF-IDF)

...	rule/VBN	...	rule/NNS	...	out/IN	...	consideration/NN	...
...	0.37	...	0.32	...	0.04	...	0.27	...

$tfidf(t, d, \boldsymbol{D}) = tf(t, d) \times idf(t, \boldsymbol{D})$ where, e.g.,

$$tf(t, d) = 0.5 + \frac{0.5 \times f(t,d)}{max\{f(w,d) : w \in d\}}$$

$$idf(t, \boldsymbol{D}) = log \frac{|\boldsymbol{D}|}{|\{t \in \boldsymbol{D} : t \in d\}|}$$

Presentation Overview

Intuition
- Natural Language Processing
- Machine Learning
- Natural Language Processing and Machine Learning

Natural Language Processing for Machine Learning

Machine Learning for Natural Language Processing

Sentence Classification

Each numeric row vector representing a document can be understood as $\boldsymbol{x}_i$.

Collection of all $\boldsymbol{x}_i$ becomes $\boldsymbol{X}$ (features).

Suppose that we know to which category each document belongs (news, sport, science, culture).

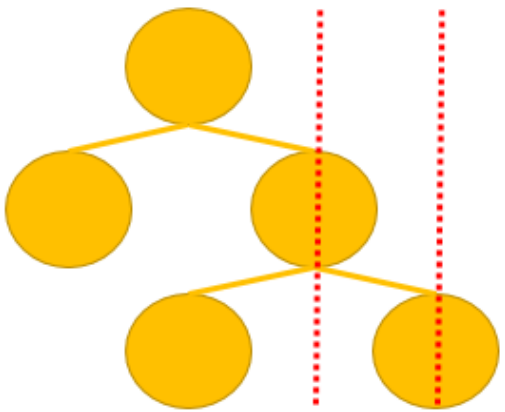
We can represent categories numerically and order them appropriately in column vector $\boldsymbol{y}$ (labels).

We get $\boldsymbol{D} = \langle \boldsymbol{X}, \boldsymbol{y} \rangle$ (dataset).

Sentence Classification (cont.)

We divide $\boldsymbol{D}$ into $\boldsymbol{D_{train}}$ and $\boldsymbol{D_{test}}$, e.g., we randomly sample each $\boldsymbol{x_i}$ to $\boldsymbol{D_{test}}$ with probability 0.2. Remaining $\boldsymbol{x_i}$ become $\boldsymbol{D_{train}}$.

We use $\boldsymbol{D_{train}}$ to train classification function $f(\boldsymbol{x_i})$ (e.g., decision tree classifier).



Now we have $f(\boldsymbol{x}_i)$ which we can use to assign categories to documents represented in the same semantic space.

Sentence Classification (cont.)

We use $\boldsymbol{D_{test}}$ to evaluate performance of our classifier.

Accuracy is a simple evaluation metric we can use. It is computed as ratio of correctly classified $\boldsymbol{x}_i \in \boldsymbol{D_{train}}$ over all $\boldsymbol{x}_i \in \boldsymbol{D_{train}}$.

If we want more insight into behaviour of our classifier we can use $\boldsymbol{D_{test}}$ to produce confusion matrix.

	1	2	3
1	24	0	3
2	2	12	1
3	5	0	52

References I


Bishop Christopher. 2006. *Pattern Recognition and Machine Learning*. Springer.


Jurafsky Daniel, and James H. Martin. 2009. *Speech and Language Processing: An Introduction to Natural Language Processing, Speech Recognition, and Computational Linguistics*. 2nd edition. Prentice-Hall.


Manning Christopher D. and Hinrich Schütze. 1999. *Foundations of Statistical Natural Language Processing*. The MIT Press.


Turney Peter D. and Patrick Pantel. From frequency to meaning: Vector space models of semantics. *Journal of artificial intelligence research*, 2010, 37.1: 141-188.


Hauskrecht Milos. *CS2750 Machine Learning* [online]. 2015 [cit. 06/10/2015]. Accessed at: `https://people.cs.pitt.edu/~milos/courses/cs2750/`


Litman Diane. *CS2731 Introduction to Natural Language Processing* [online]. 2013 [cit. 06/10/2015]. Accessed at: `http://people.cs.pitt.edu/~litman/courses/cs2731/2731.html`

Thank you!

Questions, comments and suggestions are welcome now
or any time at jas438@pitt.edu.