Natural Language Understanding

Lecture 1: Introduction

Adam Lopez
TAs: Marco Damonte, Federico Fancellu, Ida Szubert, Clara Vania
Credits: much material by Mirella Lapata and Frank Keller

16 January 2018

School of Informatics
University of Edinburgh
alopez@inf.ed.ac.uk

Reading: Goldberg (2015), Manning (2015)

Introduction

What is Natural Language Understanding?

Natural language understanding:

- often refers to full comprehension/semantic processing of language;
- here, natural language understanding is used to contrast with natural language generation.

Understanding:

Text $\Longrightarrow$ Analyses (parse trees, logical forms, discourse segmentation, etc.)

Generation:

Non-linguistic input (logical forms, database entries, etc.) or text $\Longrightarrow$ Text

Course Content

NLU covers advanced NLP methods, with a focus on *learning representations,* at all levels: words, syntax, semantics, discourse.

We will focus on *probabilistic models* that use *deep learning methods* covering:

- word embeddings;
- feed-forward neural networks;
- recurrent neural networks;
- (maybe) convolutional neural networks.

We will also touch on discriminative and unsupervised learning.

Course Content

Deep architectures and algorithms will be applied to NLP tasks:

- language modeling
- part-of-speech tagging
- syntactic parsing
- semantic parsing
- (probably) sentiment analysis
- (probably) discourse coherence
- (possibly) other things

The assignments will involve practical work with deep models.

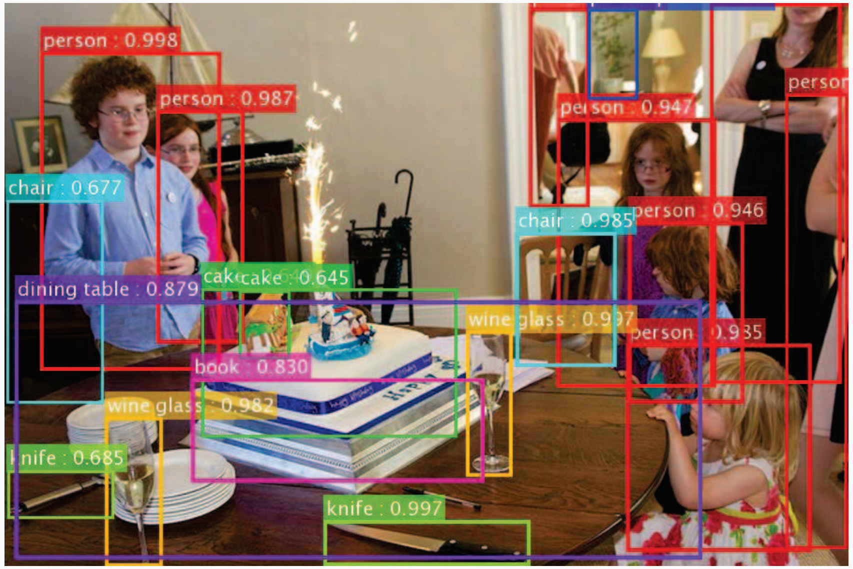
Why Deep Learning?

The Success of Deep Models: Speech Recognition

Deep belief networks (DBNs) achieve a 33% reduction in word error rate (WER) over an HMM with Gaussian mixture model (GMM) (**?**):

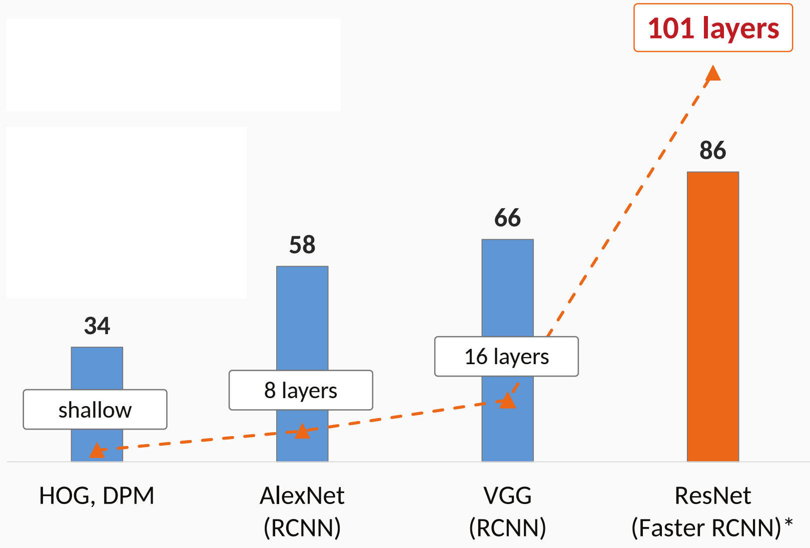
		WER	
MODELING TECHNIQUE	**#PARAMS [10^6]**	**HUB5'00-SWB**	**RT03S-FSH**
GMM, 40 MIX DT 309H SI	29.4	23.6	27.4
NN 1 HIDDEN-LAYER × 4,634 UNITS	43.6	26.0	29.4
+ 2 × 5 NEIGHBORING FRAMES	45.1	22.4	25.7
DBN-DNN 7 HIDDEN LAYERS × 2,048 UNITS	45.1	17.1	19.6
+ UPDATED STATE ALIGNMENT	45.1	16.4	18.6
+ SPARSIFICATION	15.2 NZ	16.1	18.5
GMM 72 MIX DT 2000H SA	102.4	17.1	18.6

The Success of Deep Models: Object Detection



Source: Kaiming He: Deep Residual Learning: MSRA @ ILSVRC & COCO 2015 competitions. Slides.

The Success of Deep Models: Object Detection

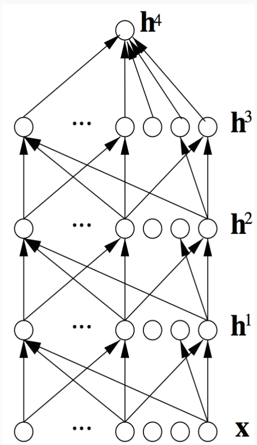


PASCAL VOC 2007 **Object Detection** mAP (%)

Source: Kaiming He: Deep Residual Learning: MSRA @ ILSVRC & COCO 2015 competitions. Slides.

Representation Learning

Why do deep models work so well (for speech and vision at least)?
Because they are good at *representation learning:*



Source: Richard Socher: Introduction to CS224d. Slides.

Neural nets learn multiple representations $\mathbf{h}^n$ from an input $\mathbf{x}$.

Representation Learning vs. Feature Engineering

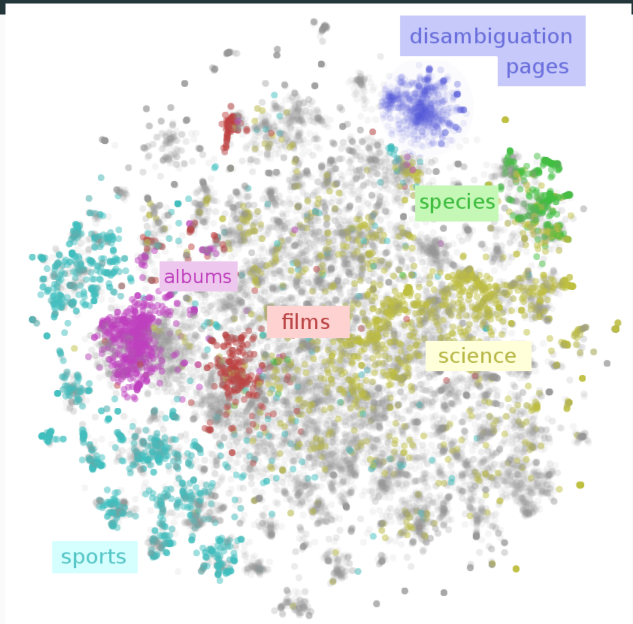
What's the appeal of representation learning?

- manually designed features are over-specified, incomplete and take a long time to design and validate;
- learned representations are easy to adapt, fast to obtain;
- deep learning provides a very flexible, trainable framework for representing world, visual, and linguistic information;
- in probabilistic models, deep learning frees us from having to make independence assumptions.

In short: deep learning solves many things that are difficult about machine learning... rather than NLP, which is still difficult!

Adapted from Richard Socher: Introduction to CS224d. Slides.

Representation Learning: Words



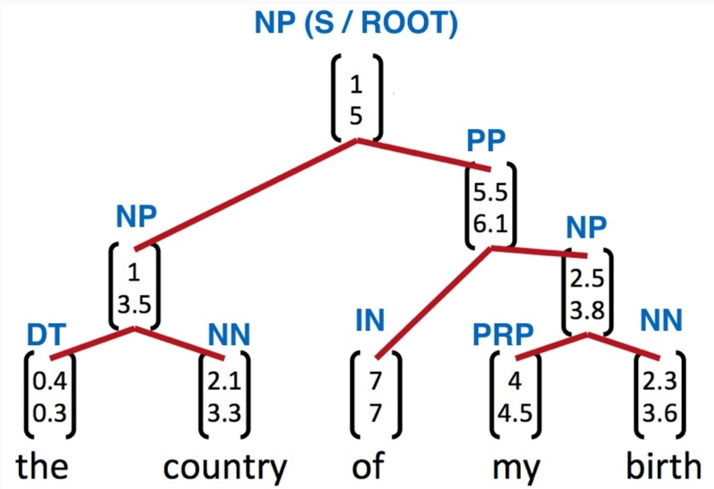
Source: http://colah.github.io/posts/2015-01-Visualizing-Represent

The figure shows an example of learned representations: word embeddings. These are just vectors in $\mathbb{R}^n$, where n is typically something like 300, 500, or 1000.

They're difficult to interpret, but if you project them onto a two-dimensional space (as in the figure) and visualize the space, as we've done with this wikipedia data, then you'll tend to see interesting patterns. In this example, we've highlighted words from the same category in wikipedia.

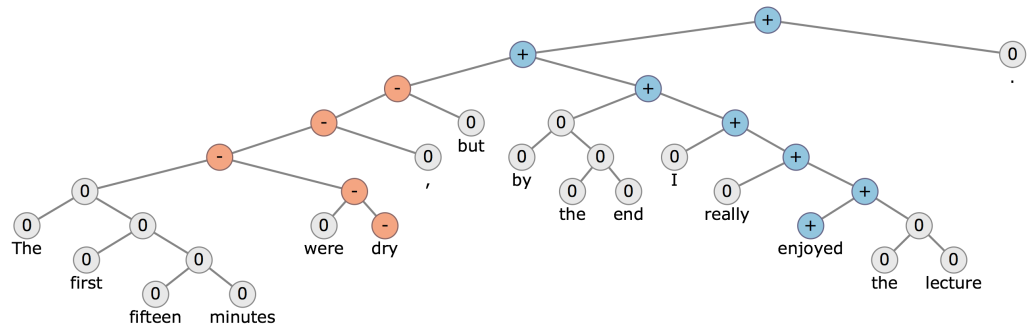
Intuitively, these representations capture semantic similarity. Though this intuition can be taken too far.

Representation Learning: Syntax



Source: Roelof Pieters: Deep Learning for NLP: An Introduction to Neural Word Embeddings. Slides.

Representation Learning: Sentiment



Source: Richard Socher: Introduction to CS224d. Slides.

Supervised vs. Unsupervised Methods

Standard NLP systems use a supervised paradigm:

Training:

Labeled training data $\Longrightarrow$ Features, representations $\Longrightarrow$ Prediction procedure (trained model)

Supervised vs. Unsupervised Methods

Standard NLP systems use a supervised paradigm:

Testing:

Unlabeled test data $\Longrightarrow$ Features, representations $\Longrightarrow$ Prediction procedure (from training) $\Longrightarrow$ Labeled output

Supervised vs. Unsupervised Methods

NLP has often focused on *unsupervised learning,* i.e., learning without labeled training data:

Unlabeled data $\Longrightarrow$ Features, representations $\Nearrow$ Clustered output

Unlabeled data $\Longrightarrow$ Features, representations $\Searrow$ Prediction procedure

Deep models can be employed both in a supervised and an unsupervised way. Can also be used for *transfer learning*, where representations learned for one problem are reused in another.

Supervised vs. Unsupervised Methods

Example of unsupervised task we'll cover:

Part of speech induction:

walk	$\Longrightarrow$	walk.VVB
runners		runners.NNS
keyboard		keyboard.NN
desalinated		desalinate.VVD

Course Mechanics

Relationship to other Courses

Natural Language Understanding:

- requires: Accelerated Natural Language Processing OR Informatics 2A and Foundations of Natural Language Processing;
- complements: Machine Translation; Topics in Natural Language Processing.

Machine learning and programming:

- IAML, MLPR, or MLP (can be taken concurrently);
- CPSLP or equivalent programming experience.

A few topics may also be covered in MLP or MT.

Background

Background required for the course:

- You should be familiar with Jurafsky and Martin (2009)
- But this textbook serves as background only. Each lecture will rely on one or two papers as the main reading. The readings are assessible: read them and discuss.
- You will need solid maths: probability theory, linear algebra, some calculus.
- for a maths revision, see Goldwater (2015).

Course Mechanics

- NLU will have 15 lectures, 1 guest lecture, 2 feedforward sessions; no lectures in flexible learning week;
- `http://www.inf.ed.ac.uk/teaching/courses/nlu/`
- see course page for lecture slides, lecture recordings, and materials for assignments;
- course mailing list: `nlu-students@inf.ed.ac.uk`; you need to enroll for the course to be subscribed;
- the course has a Piazza forum; use it to discuss course materials, assignments, etc.;
- assignments will be submitted using TurnItIn (with plagiarism detection) on Learn;
- *You need a DICE account!* If you dont have one, apply for one through the ITO as soon as possible.

Assessment

Assessment will consist of:

- one assessed coursework, worth 30%. Pair work is strongly encouraged.
- a final exam (120 minutes), worth 70%.

Key dates:

- Assignment issued week 3.
- Assigment due March 8 at 3pm (week 7).
- Assignment will include intermediate milestones and a suggested timeline.

Assignment deadline will be preceded by *feedforward sessions* in which you can ask questions about the assignment.

Feedback

Feedback students will receive in this course:

- the course includes short, non-assessed quizzes;
- these consist of multiple choice questions and are marked automatically;
- each assignment is preceded by a feedforward session in which students can ask questions about the assignment;
- the discussion forum is another way to get help with the assignments; it will be monitored once a day by course staff;
- the assignment will be marked within two weeks;
- individual, written comments will be provided by the markers and sample solutions will be released.

How to get help

Ask questions. Asking questions is how you learn.

- In-person office hour (starting week 3). Details TBA.
- *Virtual* office hour (starting week 3). Details TBA.
- piazza forum: course staff will answer questions once a day, Monday through Friday. You can answer questions any time! Your questions can be private, and/ or anonymous to classmates.
- Don't ask me questions over email. I might not see your question for days. And when I do, I will just repost it to piazza.