Natural Language Understanding

Lecture 1: Introduction

Frank Keller and Mirella Lapata

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

January 17, 2017

Reading: Goldberg (2015); Manning (2015).

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.)

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: lexicon, syntax, semantics, discourse.

We will introduce *deep learning methods,* covering:

- word embeddings;
- feed-forward neural networks;
- recurrent neural networks;
- recursive neural networks;
- convolutional neural networks.

We will also compare deep learning models with conventional discriminative and unsupervised learning models.

Course Content

Deep architectures and algorithms will be applied to NLP tasks:

- language modeling;
- part-of-speech tagging;
- parsing;
- semantic role labeling;
- semantic composition;
- sentiment analysis;
- discourse coherence.

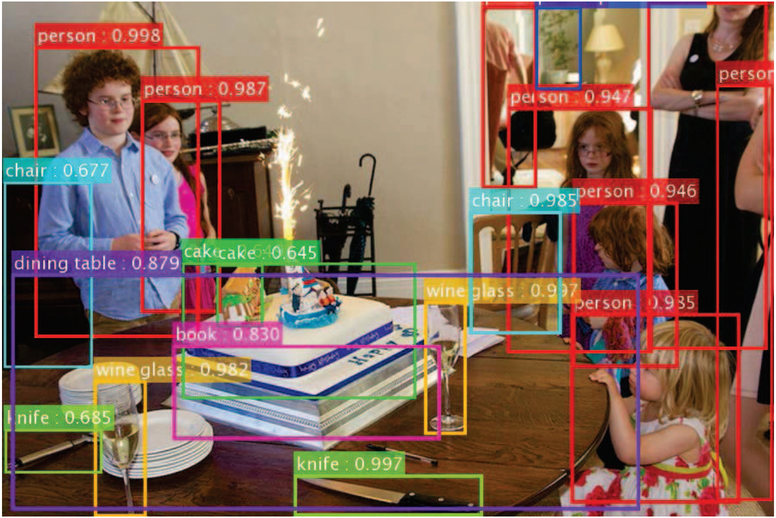
The assignments will involve practical work with deep models.

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) (Hinton et al. 2012):

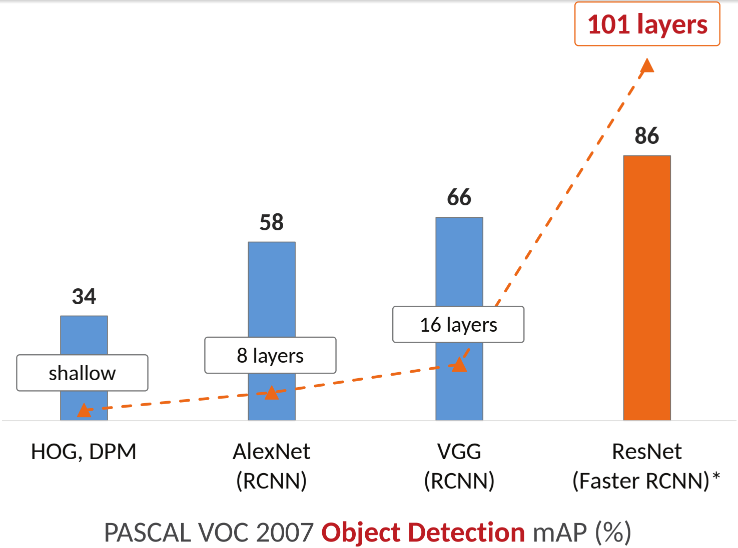
		WER	
MODELING TECHNIQUE	**#PARAMS [10^6]**	**HUB5'00-SWB**	**RT03S-FSH**
GMM, 40 MIX DT 309H SI	29.4	23.6	27.4
NN 1 HIDDEN-LAYER × 4,634 UNITS	43.6	26.0	29.4
+ 2 × 5 NEIGHBORING FRAMES	45.1	22.4	25.7
DBN-DNN 7 HIDDEN LAYERS × 2,048 UNITS	45.1	17.1	19.6
+ UPDATED STATE ALIGNMENT	45.1	16.4	18.6
+ SPARSIFICATION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

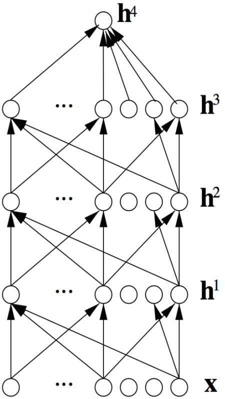
The Success of Deep Models: Object Detection



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 $\mathbf{x}$.

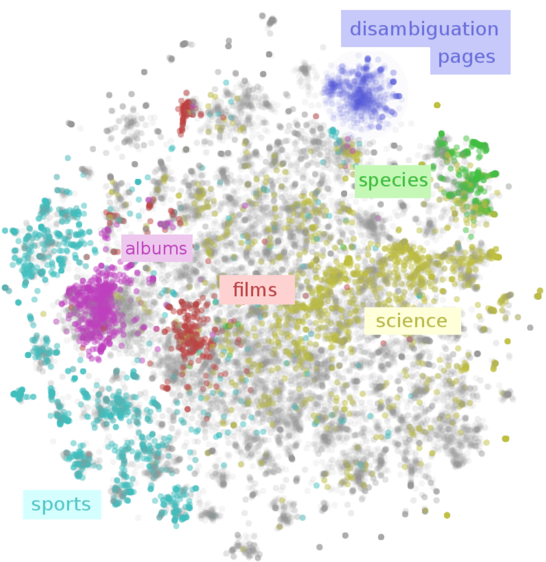
Representation Learning vs. Feature Engineering

Reasons for exploring deep 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;
- deep learning can learn be unsupervised (from raw text) or supervised (with specific labels like positive/negative).

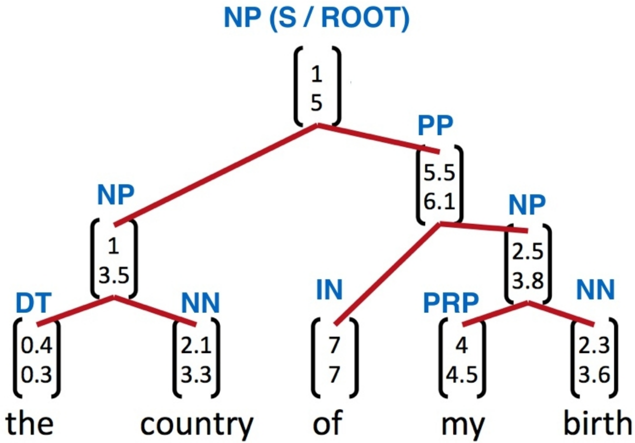
Source: Richard Socher: Introduction to CS224d. Slides.

Representation Learning: Lexicon



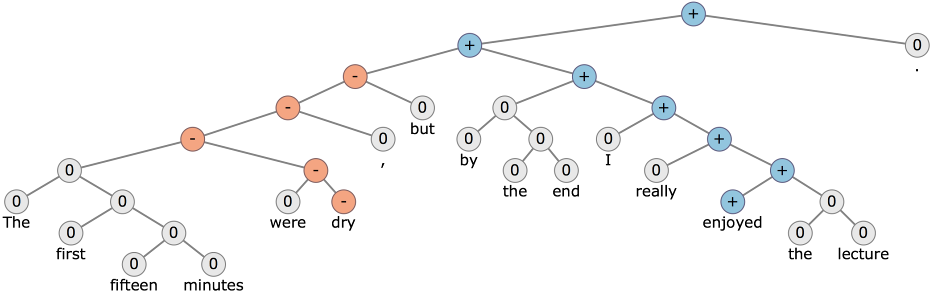
Source: http://colah.github.io/posts/2015-01-Visualizing-Representations/

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

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

Unlabeled data $\Longrightarrow$ Features, representations $\nearrow$ Labeled output / $\searrow$ Prediction procedure

Deep models can be employed both in a supervised and an unsupervised way.

Supervised vs. Unsupervised Methods

Some unsupervised tasks we'll cover:

Topic models:

⟹

Topic 5

word	prob.
RED	.202
BLUE	.099
GREEN	.096
YELLOW	.073
WHITE	.048
COLOR	.048
BRIGHT	.030
COLORS	.029
ORANGE	.027
BROWN	.027
PINK	.017
LOOK	.017
BLACK	.016
PURPLE	.015
CROSS	.011
COLORED	.009

Topic 43

word	prob.
MIND	.081
THOUGHT	.066
REMEMBER	.064
MEMORY	.037
THINKING	.030
PROFESSOR	.028
FELT	.025
REMEMBERED	.022
THOUGHTS	.020
FORGOTTEN	.020
MOMENT	.020
THINK	.019
THING	.016
WONDER	.014
FORGET	.012
RECALL	.012

Topic 56

word	prob.
DOCTOR	.074
DR.	.063
PATIENT	.061
HOSPITAL	.049
CARE	.046
MEDICAL	.042
NURSE	.031
PATIENTS	.029
DOCTORS	.028
HEALTH	.025
MEDICINE	.017
NURSING	.017
DENTAL	.015
NURSES	.013
PHYSICIAN	.012
HOSPITALS	.011

Images: `http://www.progressarkansas.com/news.htm`, Steyvers and Griffiths (2007).

Supervised vs. Unsupervised Methods

Some unsupervised tasks we'll cover:

Part of speech induction:

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

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.

Additional prerequisites:

- IAML or MLPR;
- CPSLP or equivalent programming experience.

Some overlap between NLU and MLP.

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;
- you will need solid maths: probability theory, linear algebra, some calculus;
- for a maths revision, see Goldwater (2015).

Course Mechanics

- NLU runs weeks 1–10, with 19 slots: 17 lectures, 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:

- two assessed assignments, worth 15% each (30% in total);
- a final exam (120 minutes), worth 70%.

Due dates:

- Assignment 1 (Word Embeddings): issued Feb 3, due Feb 17;
- Assignment 2 (Language Modeling): issued Mar 14, due Mar 28.

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

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 by the lecturers and TAs;
- the assignment will be marked within two weeks;
- individual, written comments will be provided by the markers and sample solutions will be released.

References

Goldberg, Yoav. 2015. A primer on neural network models for natural language processing. Unpublished mansucript, arXiv:1510.00726.

Goldwater, Sharon. 2015. Mathematics tutorials. University of Edinburgh, `http://homepages.inf.ed.ac.uk/sgwater/math_tutorials.html`.

Hinton, Geoffrey, Li Deng, Dong Yu, George E. Dahl, Abdel-rahman Mohamed, Navdeep Jaitly, Andrew Senior, Vincent Vanhoucke, Patrick Nguyen, Tara N. Sainath, and Brian Kingsbury. 2012. Deep neural networks for acoustic modeling in speech recognition. *IEEE Signal Processing Magazine* 29(6): 82–97.

Jurafsky, Daniel, and James H. Martin. 2009. *Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics and Speech Recognition*. Upper Saddle River, NJ: Pearson Education, 2nd edn.

Manning, Christopher D. 2015. Computational linguistics and deep learning. *Computational Linguistics* 41(4): 701–707.

Steyvers, M., and T. Griffiths. 2007. Probabilistic topic models. In T. Landauer, D. McNamara, S. Dennis, and W. Kintsch, eds., *Latent Semantic Analysis: A Road to Meaning*. Hillsdale, NJ: Lawrence Erlbaum Associates.