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
Introduction

What is Natural Language Understanding?

Course Content

Why Deep Learning?

The Success of Deep Models

Representation Learning

Unsupervised Models

Course Mechanics

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 $\rightarrow$ 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:

\[
\text{Text} \quad \rightarrow \quad \text{Analyses (parse trees, logical forms, discourse segmentation, etc.)}
\]

Generation:

\[
\text{Non-linguistic input (logical forms, database entries, etc.) or text} \quad \rightarrow \quad \text{Text}
\]
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.
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) (?):

<table>
<thead>
<tr>
<th>MODELING TECHNIQUE</th>
<th>#PARAMS $[10^6]$</th>
<th>HUB5’00-SWB</th>
<th>RT03S-FSH</th>
</tr>
</thead>
<tbody>
<tr>
<td>GMM, 40 MIX DT 309H SI</td>
<td>29.4</td>
<td>23.6</td>
<td>27.4</td>
</tr>
<tr>
<td>NN 1 HIDDEN-LAYER $\times$ 4,634 UNITS</td>
<td>43.6</td>
<td>26.0</td>
<td>29.4</td>
</tr>
<tr>
<td>+ 2 $\times$ 5 NEIGHBORING FRAMES</td>
<td>45.1</td>
<td>22.4</td>
<td>25.7</td>
</tr>
<tr>
<td>DBN-DNN 7 HIDDEN LAYERS $\times$ 2,048 UNITS</td>
<td>45.1</td>
<td>17.1</td>
<td>19.6</td>
</tr>
<tr>
<td>+ UPDATED STATE ALIGNMENT</td>
<td>45.1</td>
<td>16.4</td>
<td>18.6</td>
</tr>
<tr>
<td>+ SPARSIFICATION</td>
<td>15.2 NZ</td>
<td>16.1</td>
<td>18.5</td>
</tr>
<tr>
<td>GMM 72 MIX DT 2000H SA</td>
<td>102.4</td>
<td>17.1</td>
<td>18.6</td>
</tr>
</tbody>
</table>
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

HOG, DPM 34 shallow

AlexNet (RCNN) 58 8 layers

VGG (RCNN) 66 16 layers

ResNet (Faster RCNN)* 86 101 layers

PASCAL VOC 2007 **Object Detection** mAP (%)
Why do deep models work so well (for speech and vision at least)? Because they are good at *representation learning*:

Neural nets learn multiple representations $h^n$ from an input $x$. 
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

Standard NLP systems use a supervised paradigm:

**Training:**

Labeled training data $\rightarrow$ Features, representations $\rightarrow$ Prediction procedure (trained model)
Supervised vs. Unsupervised Methods

Standard NLP systems use a supervised paradigm:

**Testing:**

Unlabeled test data $\rightarrow$ Features, representations $\rightarrow$ Prediction procedure (from training) $\rightarrow$ Labeled output
NLP has often focused on *unsupervised learning*, i.e., learning without labeled training data:

Unlabeled data $\rightarrow$ Features, representations $\rightarrow$ Clusters $\rightarrow$ 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.
Example of unsupervised task we’ll cover:

Part of speech induction:

walk \rightarrow walk.VVB
runners \rightarrow runners.NNS
keyboard \rightarrow keyboard.NN
desalinated \rightarrow 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 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/](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 don’t 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.
- Assignment 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 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.