Natural Language Processing: Part II Overview of Natural Language Processing (L90): ACS Lecture 12

Ann Copestake

Computer Laboratory
University of Cambridge

November 2017

Recent NLP research

Visual question answering

Shapeworld

DELPH-IN/English Resource Grammar

Where is NLP going?

Outline.

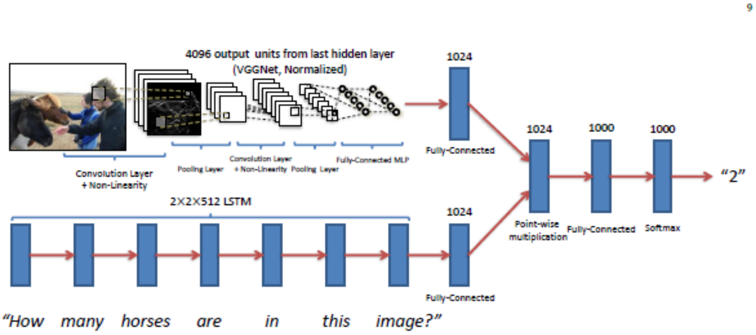
Multimodal architectures

- Input to a NN is just a vector: we can combine vectors from different sources.
- e.g., features from a CNN for visual recognition concatenated with word embeddings.
- multimodal systems: captioning, visual question answering (VQA).

Visual Question Answering

- System is given a picture and a question about the picture which it has to answer.
- Best known dataset: COCO VQA (Agrawal et al, 2016).
- Questions and answers for images from Amazon Mechanical Turk.
- Task: provide questions which humans can easily answer but can "stump the smart robot" (cf Turing Test!)
- Three questions per image.
- Answers from 10 different people.
- Also asked for answers without seeing the image (22%).

VQA architecture (Agrawal et al, 2016)



Why does this male have his arms in this position?	balance for balance for balance	angry he's carrying bags hug
Are the clouds high in the sky?	yes yes yes	no no yes

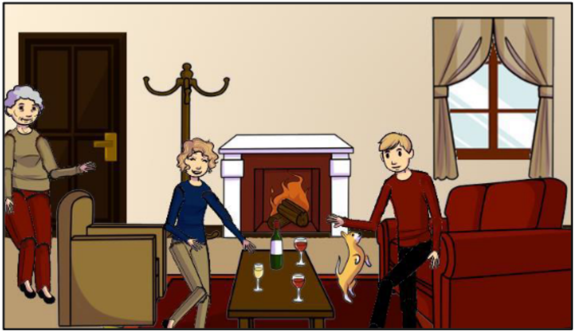
from Agrawal et al (2016)

Which player on the field head-butted the ball?	18 18 player on left	1 in front of goal number 13 number 22
What number is on the girl in black?	18 18 18	1 4 8

from Agrawal et al (2016)

Is this person trying to hit a ball?	yes yes yes	yes yes yes
What is the person hitting the ball with?	frisbie racket round paddle	bat bat racket

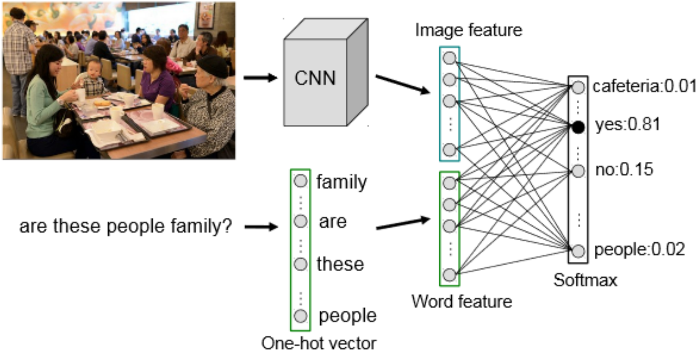
from Agrawal et al (2016)



How many glasses are on the table?	3 3 3	2 2 6
What is the woman reaching for?	door handle glass wine	fruit glass remote

from Agrawal et al (2016)

Baseline system



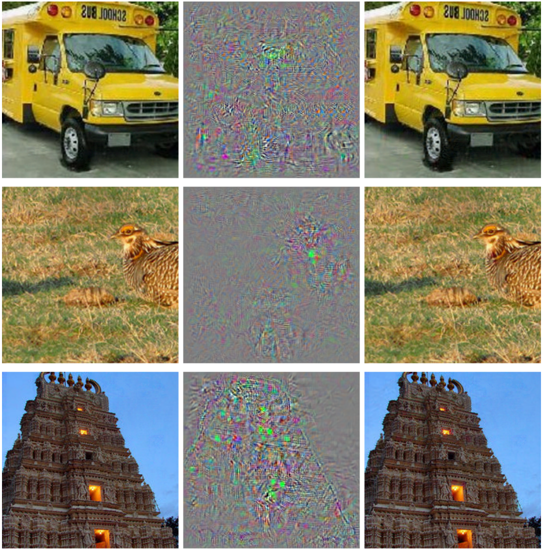
http://visualqa.csail.mit.edu/

Learning commonsense knowledge?

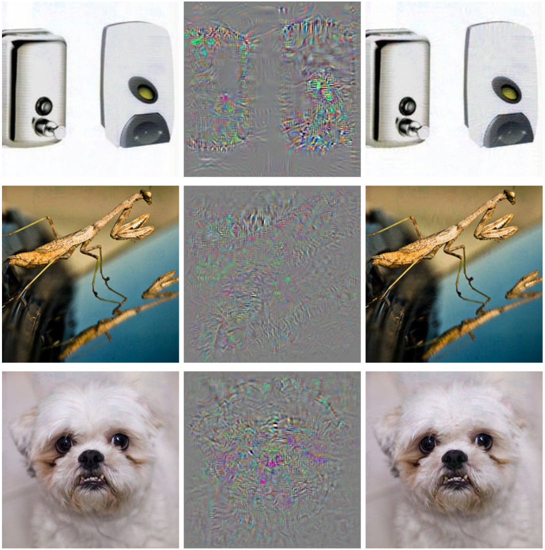
- Zhou et al's baseline system (no hidden layers) performs as well as systems with much more complex architectures (55.7%).
- Correlates input words and visual concepts with the answer.
- Systems are much better than humans at answering without seeing the image (BOW model is at 48%).
- To an extent, the systems are discovering biases in the dataset.
- Systems make errors no human would ever make on unexpected questions: e.g., 'Is there an aardvark?'

Adversarial examples

- For image recognition, images that are correctly recognised are perturbed in a manner imperceptible to a human and are then not recognised.
  https://arxiv.org/pdf/1312.6199.pdf
- Systematically find adversarial examples via low probability ‘pockets’ (because the space is not smooth): these can’t be found efficiently by random sampling around a given example.
- Not clear whether anything directly comparable for NLP: though https://arxiv.org/pdf/1707.07328.pdf for reading comprehension.
- also ‘Build it, Break it: the language edition’
  https://bibinlp.umiacs.umd.edu/



https://arxiv.org/pdf/1312.6199.pdf



https://arxiv.org/pdf/1312.6199.pdf

Outline.

Visual question answering

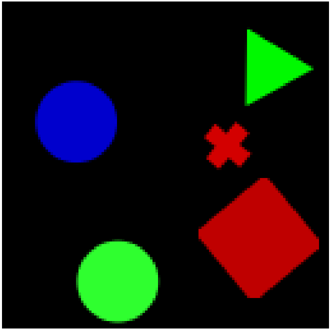
Shapeworld

DELPH-IN/English Resource Grammar

Where is NLP going?

Shapeworld (Kuhnle and Copestake, 2016, 2017)

Training and testing NNs with grounded language:



All circles are to the left of a red cross.

$$\forall s_1 \in W: \text{circle}(s_1.\text{shape}) \Rightarrow$$
$$\left(\exists s_2 \in W: \text{cross}(s_2.\text{shape}) \wedge \text{red}(s_2.\text{colour}) \wedge s_1.x < s_2.x\right)$$

Shapeworld (cont.)

- Automatically generate huge number of models in various classes: generate diagrams and DMRS using models.
- Generate English captions from DMRS using ERG/ACE (both true and false captions).
- Use pictures and captions to train NNs: evaluate performance on examples including unseen combinations (e.g., red triangle).
- Finding: performance of successful standard VQA approaches surprisingly bad (need new models).
- In progress: more languages.
- Compared with alternatives: no need for human annotation, less limited than simple template generation.

Outline.

Visual question answering

Shapeworld

DELPH-IN/English Resource Grammar

Where is NLP going?

English Resource Grammar

ERG is a symbolic, bidirectional, hand-written grammar of English, begun in the 1990s, mainly written by Dan Flickinger. Subsequent development of grammars for other languages.

ERG demos:
Michael Goodman:
`http://chimpanzee.ling.washington.edu/demophin`
Ned Letcher:
`http://delph-in.github.io/delphin-viz/demo/`
Both running the ERG (and other grammars) using Woodley Packard's ACE parser/generator.

What can one do with the ERG?

Applications investigated include:

- Machine translation: e.g., Bond et al (2011)
- Information extraction and QA: e.g., MacKinlay et al (2009)
- Ontology extraction: e.g., Herbelot and Copestake (2006)
- Question generation: e.g., Yao et al (2012)
- Entailment recognition: e.g., Lien and Kouylekov (2014)
- Preprocessing for distributional semantics: e.g., Herbelot (2013), Emerson and Copestake (2016)
- Detection scope of negation: e.g., Packard, Bender, Read, Oepen and Dridan (2014)
- Robot control interface: e.g., Packard (2014)
- Language teaching: Suppes, Flickinger et al (2012)
- Logic to English (for teaching logic): e.g. Flickinger (2016)
- Proposition based summarization (Fang et al, 2016)

ERG in practical use

- Dan Flickinger: *Language Arts and Writing for Grades 2–6*
- Automated system for children to practice writing English.
- Prompt: *Ricky stores his toys in the closet. Where are Ricky's toys?*
- Child's task: create a grammatically correct answer from a set of words provided (organised by part of speech).
- System's task: check that the answer is grammatically correct and plausible,
- and, if possible, provide feedback on mistakes.

Language learning examples

Some correct answers:

Ricky’s toys are in the closet
Ricky’s toys are in his closet
they are in his closet
the toys are in Ricky’s closet
Ricky’s toys are in Ricky’s closet

Some incorrect answers:

Ricky’s are his toys in his closet
in in in in Ricky’s in Ricky’s Ricky’s
Ricky’s toys are in the garden

Role of ERG

- Precision grammar (modified ERG) needed for checking, broad coverage because lots of exercises.
- Compositional semantics to check syntactically correct answer is plausible.
- 20,000 students creating 500,000 sentences per week

Grade 5 example:
Abigail didn't want to go hiking with her parents because she felt too tired and wanted to rest instead. Why didn't Abigail want to go hiking?

Role of ERG

- Precision grammar (modified ERG) needed for checking, broad coverage because lots of exercises.
- Compositional semantics to check syntactically correct answer is plausible.
- 20,000 students creating 500,000 sentences per week

Grade 5 example:
Abigail didn't want to go hiking with her parents because she felt too tired and wanted to rest instead. Why didn't Abigail want to go hiking?

Role of ERG

- Precision grammar (modified ERG) needed for checking, broad coverage because lots of exercises.
- Compositional semantics to check syntactically correct answer is plausible.
- 20,000 students creating 500,000 sentences per week

Grade 5 example:
Abigail didn't want to go hiking with her parents because she felt too tired and wanted to rest instead. Why didn't Abigail want to go hiking?

Outline.

Visual question answering

Shapeworld

DELPH-IN/English Resource Grammar

Where is NLP going?

Deep learning methodology?

- Methodological issues are fundamental to deep learning: e.g., subtle biases in training data will be picked up.
- Old tasks and old data, suitable for older styles of machine learning, are possibly no longer appropriate.
- The lack of predefined interpretation of the latent variables is what makes the models more flexible/powerful ...
- but the models are usually not interpretable by humans after training: serious practical and ethical issues.
- Questioning of very fundamental assumptions: e.g., Hinton on backprop: "My view is throw it all away and start again".

Deep learning in NLP

- Much current interest in micro-worlds and ‘small data’.
- Linguistics may be regaining importance:
  - good Bayesians should properly investigate priors . . .
  - interest in what systems can do in principle
  - building proper tests and evaluations

Please beware of hype!

> *What can you say to a software package that speaks English better than the average gas-station attendant?*

Unix World article, Alan Southerton, May 1989, about the ‘Natural Language’ interface system

Deep learning in NLP

- Much current interest in micro-worlds and ‘small data’.
- Linguistics may be regaining importance:
  - good Bayesians should properly investigate priors . . .
  - interest in what systems can do in principle
  - building proper tests and evaluations

Please beware of hype!

> *What can you say to a software package that speaks English better than the average gas-station attendant?*

Unix World article, Alan Southerton, May 1989, about the ‘Natural Language’ interface system

Deep learning in NLP

- Much current interest in micro-worlds and 'small data'.
- Linguistics may be regaining importance:
  - good Bayesians should properly investigate priors . . .
  - interest in what systems can do in principle
  - building proper tests and evaluations

Please beware of hype!

> *What can you say to a software package that speaks English better than the average gas-station attendant?*

Unix World article, Alan Southerton, May 1989, about the 'Natural Language' interface system

Deep learning in NLP

- Much current interest in micro-worlds and ‘small data’.
- Linguistics may be regaining importance:
  - good Bayesians should properly investigate priors . . .
  - interest in what systems can do in principle
  - building proper tests and evaluations

Please beware of hype!

> *What can you say to a software package that speaks English better than the average gas-station attendant?*

Unix World article, Alan Southerton, May 1989, about the ‘Natural Language’ interface system

Deep learning in NLP

- Much current interest in micro-worlds and ‘small data’.
- Linguistics may be regaining importance:
  - good Bayesians should properly investigate priors . . .
  - interest in what systems can do in principle
  - building proper tests and evaluations

Please beware of hype!

> *What can you say to a software package that speaks English better than the average gas-station attendant?*

Unix World article, Alan Southerton, May 1989, about the ‘Natural Language’ interface system