L101, lecture 8

Ann Copestake

Computer Laboratory
University of Cambridge

November 2016

Outline

Outline.

Some state-of-the-art results

Making progress in distributional semantics

Machine Learning (outside NLP) and explanation

Conclusion

TOEFL: word similarity

TOEFL vocabulary test: 80 questions.

Stem:	levied
Choices:	(a) imposed
	(b) believed
	(c) requested
	(d) correlated
Solution:	(a) imposed

Human performance: non-English college entrants in US: 64%
State of the art:

TOEFL: word similarity

TOEFL vocabulary test: 80 questions.

Stem:	levied
Choices:	(a) imposed
	(b) believed
	(c) requested
	(d) correlated
Solution:	(a) imposed

Human performance: non-English college entrants in US: 64%
State of the art: 100%

State of the art: Question Answering

Jeopardy quiz show:

Prompt:

William Wilkinson's "An Account of the Principalities of Wallachia and Moldavia" inspired this author's most famous novel.

State of the art: Question Answering

Jeopardy quiz show:

Prompt:

William Wilkinson's "An Account of the Principalities of Wallachia and Moldavia" inspired this author's most famous novel.

Response:

Bram Stoker

State of the art: Question Answering

Jeopardy quiz show:

Prompt:

William Wilkinson's "An Account of the Principalities of Wallachia and Moldavia" inspired this author's most famous novel.

Response:

Bram Stoker

IBM Watson QA system beat two champion human contestants.

State of the art: Question Answering

Jeopardy quiz show:

Prompt:

William Wilkinson's "An Account of the Principalities of Wallachia and Moldavia" inspired this author's most famous novel.

Response:

Bram Stoker

IBM Watson QA system beat two champion human contestants.

“I for one welcome our new computer overlords”: Ken Jennings, a beaten contestant.

State of the art: Sentiment analysis on sentences

Sentences from movie reviews.

Even the digressions are funny.

Is this positive or negative?

State of the art: Sentiment analysis on sentences

Sentences from movie reviews.

Even the digressions are funny.

Is this positive or negative?
Positive
State of the art:

State of the art: Sentiment analysis on sentences

Sentences from movie reviews.

Even the digressions are funny.

Is this positive or negative?
Positive
State of the art: 85%

Deep Learning approach based on syntax trees by Socher et al (previous techniques 80%).

State of the art: Sentiment analysis on sentences

Sentences from movie reviews.

Even the digressions are funny.

Is this positive or negative?
Positive
State of the art: 85%

Deep Learning approach based on syntax trees by Socher et al (previous techniques 80%).

I studied at Stanford.

I studied at Harvard.

State of the art: Sentiment analysis on sentences

Sentences from movie reviews.

Even the digressions are funny.

Is this positive or negative?
Positive
State of the art: 85%

Deep Learning approach based on syntax trees by Socher et al (previous techniques 80%).

I studied at Stanford.

I studied at Harvard.

First positive, second neutral.

State of the art: paraphrase identification

Amrozi accused his brother, whom he called "the witness", of deliberately distorting his evidence.
Referring to him as only "the witness", Amrozi accused his brother of deliberately distorting his evidence.

- Standard dataset: MSRP (Microsoft Research Paraphrase Corpus): 4076 sentence pairs train, 1725 test
- All positive baseline: 66.5% (ceiling ?, not 100%)
- Best unsupervised: 74.1% accuracy (2008)
- Vector based composition, unsupervised: 73.0% (2014)
- Best supervised: distributions for sentences: 80.4% (2013)
- Best 'deep learning': 78.6% (2015)

https://www.aclweb.org/aclwiki/index.php?title=Paraphrase_Identification_(State_of_the_art)

State of the art: language comprehension tests

Text:
It was early morning. Peter Corbett helped Mark Wellman out of his wheelchair and onto the ground. They stood before El Capitan, a huge mass of rock almost three-quarters of a mile high in California's beautiful Yosemite Valley. It had been Mark's dream to climb El Capitan for as long as he could remember. But how could a person without the use of his legs hope to try to climb the highest vertical cliff on earth?
[. . .]
Question: *What had Mark Wellman long desired to do?*

1. **To accomplish one of the most difficult rock climbs in the world. (correct answer)**
2. To be the first to conquer El Capitan.
3. To climb the highest mountain in California.
4. To help his friend Peter climb El Capitan.

Martin Gleize. Textual Inference for Machine Comprehension. Computation and Language [cs.CL]. Universite Paris-Saclay, 2016.

CLEF: English and French for language learners

Results from CLEF-15 (from Gleize 2016):

System	c@1
Synapse	**0.58**
LIMSI (our system)	0.36
cicnlp	0.30
NTUNLG	0.29
CoMiC	0.29

- Baseline 0.25
- Very small dataset (87 elements test): task cannot be learned from this data.
- But a real NLU system wouldn't need to be specially trained!

State of the art: some generalisations

- Performance is often lowest on the simplest tasks for humans.
- Real broad-coverage applications have to tolerate high error rates: acceptable if high redundancy (sentiment analysis, QA) or if human intelligence can sort things out (e.g., MT).
- Most published work repeats standard tasks on standard (very limited) datasets.
- Many research evaluations bear little relationship to real tasks:
  e.g., sentiment analysis in commercial applications requires justifications.

A personal perspective

- Compared to the 1980s, really impressive performance gains on broad-coverage, shallow systems.
- Progress on classic 1980s tasks, especially inference-based, is disappointing.
- Although NLP is AI-complete, many inference tasks are designed to involve very limited world knowledge.
- Recent deep learning progress is exciting, partly because some of the main practitioners are interested in making progress on ‘deeper’ tasks.

Two topics (briefly): (i) improving distributional semantics (ii) explanation in Machine Learning.

Outline.

Distributional semantics, again

- Distributional semantics (considered broadly, including embeddings etc) is the dominant approach to lexical semantics/word meaning in NLP.
- Also cognitive approaches and (?) theoretical linguistics and (???) philosophy.
- But needs to be supplemented:
  - Combining distributions with ontologies (possibly acquired distributionally).
  - Composition (possibly using compositional semantics).
  - Contextualization (‘disambiguation’).
  - Logic/formal semantics: variety of ongoing work, including Emerson and Copestake (2016).
  - Reference/grounding: multimodal work.
  - Distributional individuals: the next few slides are all from Aurelie Herbelot.

Building distributions (A)

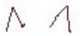
- ... saw the cat's ears twitch ...

Building distributions (A)

- ... saw the cat's ears twitch ...

Building distributions (A)

- ... saw the cat's ears twitch ...
- ... the big cat turned his head ...



Building distributions (A)

- ... saw the cat's ears twitch ...
- ... the big cat turned his head ...

Building distributions (A)

- ... saw the cat's ears twitch ...
- ... the big cat turned his head ...
- ... that the cat had dark green eyes ...

Building distributions (A)

- ... saw the cat's ears twitch ...
- ... the big cat turned his head ...
- ... that the cat had dark green eyes ...

Building distributions (A)

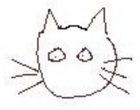
- ... saw the cat's ears twitch ...
- ... the big cat turned his head ...
- ... that the cat had dark green eyes ...
- ... paint with cat's whiskers ...

Building distributions (A)

- ... saw the cat's ears twitch ...
- ... the big cat turned his head ...
- ... that the cat had dark green eyes ...
- ... paint with cat's whiskers ...

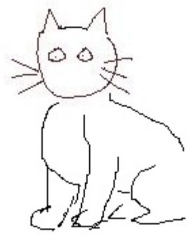
Building distributions (A)

- ... saw the cat's ears twitch ...
- ... the big cat turned his head ...
- ... that the cat had dark green eyes ...
- ... paint with cat's whiskers ...
- ... the cat stretched his legs ...



Building distributions (A)

- ... saw the cat's ears twitch ...
- ... the big cat turned his head ...
- ... that the cat had dark green eyes ...
- ... paint with cat's whiskers ...
- ... the cat stretched his legs ...



Building distributions (A)

- ... saw the cat's ears twitch ...
- ... the big cat turned his head ...
- ... that the cat had dark green eyes ...
- ... paint with cat's whiskers ...
- ... the cat stretched his legs ...
- ... the cat's enormous fluffy tail ...

Building distributions (A)

- ... saw the cat's ears twitch ...
- ... the big cat turned his head ...
- ... that the cat had dark green eyes ...
- ... paint with cat's whiskers ...
- ... the cat stretched his legs ...
- ... the cat's enormous fluffy tail ...

Building distributions (A)

- ... saw the cat's ears twitch ...
- ... the big cat turned his head ...
- ... that the cat had dark green eyes ...
- ... paint with cat's whiskers ...
- ... the cat stretched his legs ...
- ... the cat's enormous fluffy tail ...
- ... he is the cutest cat you'll ever ...

Building distributions (A)

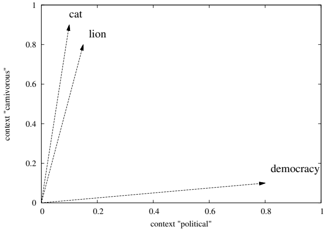
- ... saw the cat's ears twitch ...
- ... the big cat turned his head ...
- ... that the cat had dark green eyes ...
- ... paint with cat's whiskers ...
- ... the cat stretched his legs ...
- ... the cat's enormous fluffy tail ...
- ... he is the cutest cat you'll ever ...

Building distributions (A)

- ... saw the cat's ears twitch ...
- ... the big cat turned his head ...
- ... that the cat had dark green eyes ...
- ... paint with cat's whiskers ...
- ... the cat stretched his legs ...
- ... the cat's enormous fluffy tail ...
- ... he is the cutest cat you'll ever ...

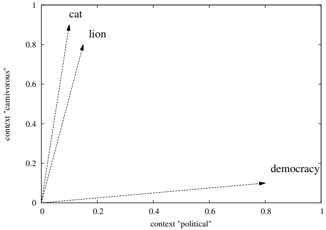
Building distributions (A)

- ... saw the cat's ears twitch ...
- ... the big cat turned his head ...
- ... that the cat had dark green eyes ...
- ... paint with cat's whiskers ...
- ... the cat stretched his legs ...
- ... the cat's enormous fluffy tail ...
- ... he is the cutest cat you'll ever ...

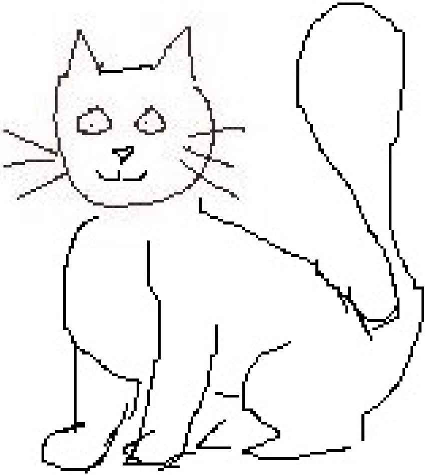


Building distributions (A)

- ... saw the cat's ears twitch ...
- ... the big cat turned his head ...
- ... that the cat had dark green eyes ...
- ... paint with cat's whiskers ...
- ... the cat stretched his legs ...
- ... the cat's enormous fluffy tail ...
- ... he is the cutest cat you'll ever ...



A distributional cat (the theory) (A)



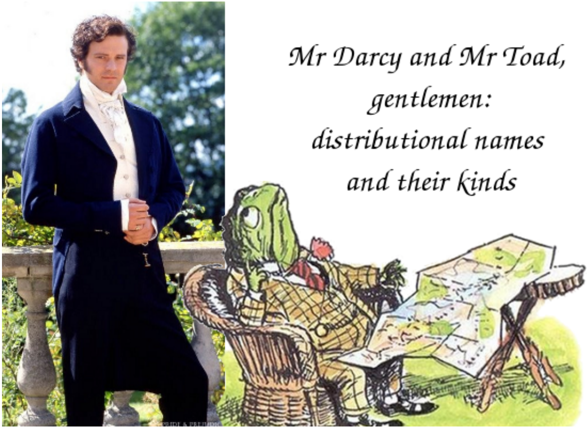
A distributional cat (the reality) (A)

Why does that cat look so bad? (A)

- Distributions model generic information.

Only 7% of NPs are references to kind

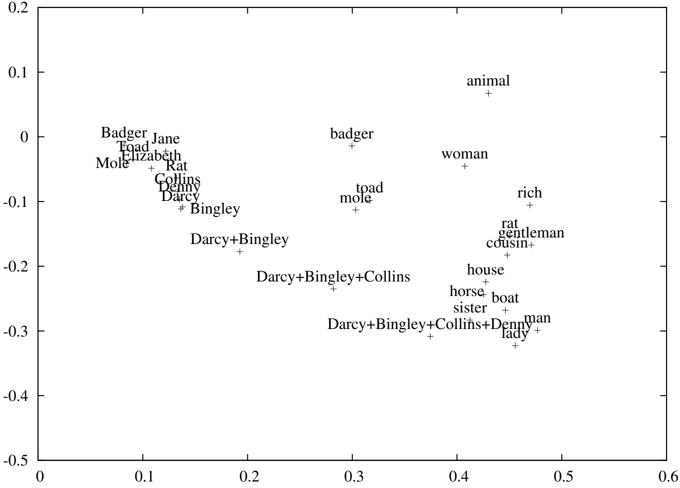
... an entirely black cat, like ...
... she owned a big ginger cat ...
... the cat was striped ...
... two long-haired white cats ...
... was a small grey cat ...
... cats are mammals ...



Aurelie Herbelot, IWCS 2015

Contextualised individuals in the BNC semantic space

(A)



Outline.

Some state-of-the-art results

Making progress in distributional semantics

Machine Learning (outside NLP) and explanation

Conclusion

Classification algorithms

- Classification is done on the basis of features extracted from data sets.
- Extracting the features (generally) requires knowledge of the domain, but actual ML classification algorithm is independent.
- Features are just numbers (or whatever): irrelevant whether blood pressure, haemoglobin count, humidity, tyre pressure ...
- Great strength (massively more scalable and robust than hand-coded rules) but also weakness.
- Models are inherently difficult to connect to language.

Interpretable models from Machine Learning

e.g., Caruana et al:

- Pneumonia risk dataset: rule-based learning system acquired:
  has asthma $\rightarrow$ lower risk
- Logistic regression deployed (though low performance) because of interpretability.
- "interpretability": users can understand the contribution of individual features in the model.
- Lou et al (2013) Generalized Additive models plus interactions: univariate plus a few pairwise interactions.

The current communication problem

Practical and legal difficulties with acceptance of ML in some applications:

1. Classifiers are only as good as their training data, but bad data values and out-of-domain input won't be recognised by a standard approach.
2. Standard classifiers cannot give any form of reason for their decisions.
3. Ideally: user could query system, system could ask for guidance, i.e., cooperative human-machine problem-solving.
4. This requires association of 'meaning' with features, and configurations of features, but also modelling co-operativity.

Computer agentivity

Decisions affecting the real world taken without human intervention:

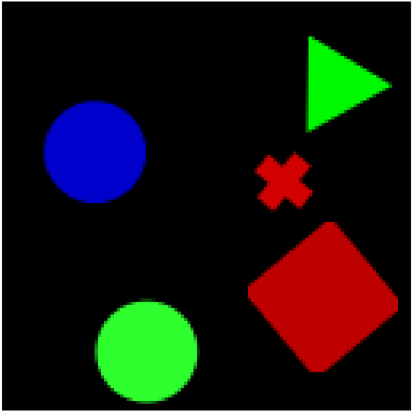
- Reaction speed: e.g., stock trading.
- Complexity of situation: e.g., load balancing (electricity grid).
- Cyber-physical systems, autonomous cars (and vacuum cleaners), internet of things.

Serious potential for harm even without Artificial General Intelligence and megalomaniac AIs.

Interpretation of features by systems

- Can the system report which features it is using?
- A form of self-reflection: associate its own state with linguistic output.
- Distributional techniques now starting to allow grounding: so symbolic input to systems possible along with perceptual data.
- Human concept formation appears to bootstrap on language:
  humans don't form concepts and then label them with language, concept formation goes along with language acquisition.

NNs with grounded quantifiers (Alex Kuhnle)



All circles are to the left of a red cross.

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

Outline.

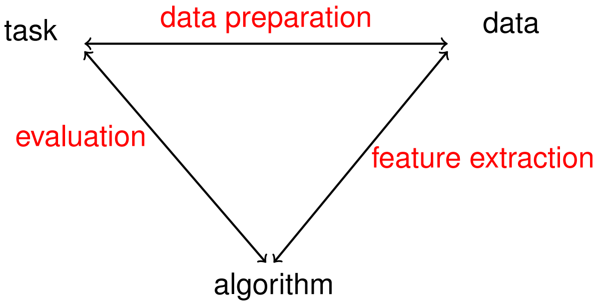
Some state-of-the-art results

Making progress in distributional semantics

Machine Learning (outside NLP) and explanation

Conclusion

Machine learning, abstractly (from lecture 1)



Standard/shared tasks (from lecture 1)