Deep Learning for Natural Language Processing

Stephen Clark
University of Cambridge and DeepMind

1. Introduction to Neural Networks for NLP

Stephen Clark
University of Cambridge and DeepMind

What's all the fuss about? (Let's look at Machine Translation)

- MT started in the 1950s
- Much harder problem than people originally thought
- Back translation output of an original system (allegedly):

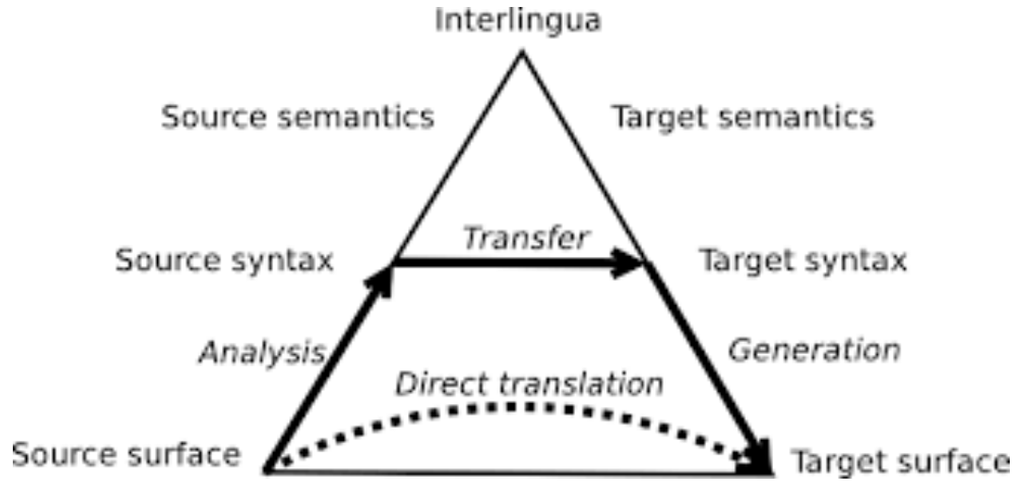
The spirit is willing but the flesh is weak ⇨ *The vodka is good but the steak is lousy*

DeepMind

UNIVERSITY OF CAMBRIDGE

Rule-Based MT and the MT Pyramid

- Rule-based MT (70s and 80s)
  - Linguist writes analysis, transfer and generation rules
  - Resulted in some working systems, e.g. SYSTRAN



DeepMind

UNIVERSITY OF CAMBRIDGE

Statistical Machine Translation (SMT)

- Started late 1990s (IBM, Jelinek) “every time I fire a linguist, my recognition rates go up”

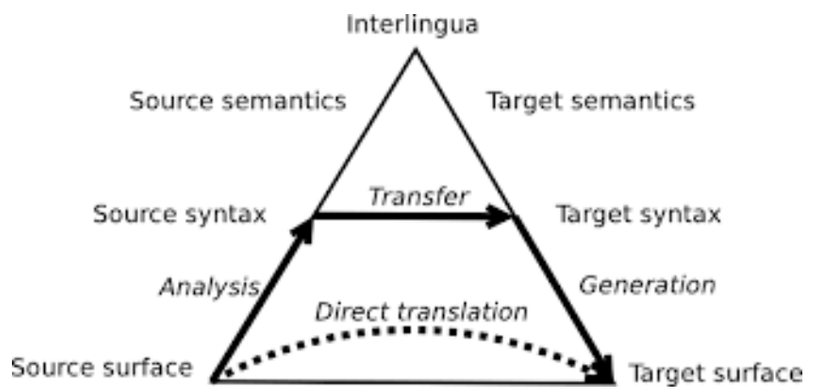
- Requires parallel corpora

- Typically no linguistic knowledge encoded

$$\arg\max_e \; p(e|f) \propto p(f|e)p(e)$$

Parallel corpus

Result	Web Result	
activities. **money laundering** and publicity campaign of the group enabling it to continue with its terrorist acts.		جملة فعالیت های مالی و پولشویی و تبلیغاتی را که زمینه و بستر فعالیت های تروریستی منافقین را فراهم می آورد ، گوشزد کرد.
tehran, 7 february 2010 (unic)— in a major step towards countering **money laundering**, the financial intelligence unit (fiu) of the islamic republic of iran and a computer based training centre (cbtc) were inaugurated today in tehran with the support of united nations office on drugs and crime (unodc).		تهران، 10 بهمن 1388 (مرکز اطلاعات سازمان ملل متحد) –در گامی بزرگ به منظور مقابله با پولشویی، واحداطلاعات مالی و مرکز آموزش رایانه ای نرم افزار مقابله با پولشویی در جمهوری اسلامی ایران با حمایت دفتر مقابله با مواد مخدر و جرم سازمان ملل متحد (unodc) در تهران افتتاح شد.
addressing the participants, mr. de leo congratulated national authorities for the establishment of the iranian fiu and for the progress made on anti-**money laundering** legislation over the last two years.		آقای دلئو ضمن تبریک به دست اندرکاران کشور برای تأسیس واحد اطلاعات مالی ایران و پیشرفت در زمینهٔ قانون مبارزه با پولشویی طی دو سال گذشته گفت
mr. de leo explained that the iranian fiu will be responsible to tackle **money laundering** and the financing of terrorism in line with international standards.		وی گفت واحد اطلاعات مالی ایران مسئول مبارزه با پولشویی و کمک مالی به تروریسم در راستای استانداردهای بین المللی است.
money laundering is the process of concealing or disguise the identity or the origin of illegal proceeds (i.e. from drug trafficking, corruption, contraband, smuggling of art and antiquities and other serious crimes) so that they appear to have originated from legitimate sources.		پولشویی روند مخفی سازی و تغییر ماهیت عمدتاً منفعت های غیر قانونی (یعنی قاچاق مواد مخدر، فساد، کالای غیر مجاز، قاچاق عتیقه و آثار هنری و سایر جرائم جدی) به نحوی است که به نظر آید این منفعت از منابع قانونی به دست آمده است.
"unodc encourages national authorities to continue progress on anti-**money laundering** and develop national legislation in countering financing of terrorism in line with international standards" mr. de leo also stated.		مسئولان کشور را تشویق می کند تا به unodc آقای دلئو همچنین گفت: " دفتر پیشرفت در زمینهٔ مبارزه با پولشویی ادامه دهند و قانون ملی مبارزه با کمک مالی به تروریسم را در راستای استانداردهای بین المللی ایجاد کنند."
"the latter, which was set up in september 2009 by the ministry of economic affairs and finance and unodc, is a centre of excellence in this country and in the sub-region capable to train officials on international best practices in tackling **money laundering** and financing of terrorism," he added.		وی افزود: "این مرکز که در ابتدا در سپتامبر 2009 (شهریور 1388) توسط وزارت امور اقتصادی و دارائی و دفتر unodc راه اندازی شده، در ایران و در سطح منطقه نمونه است و جهت آموزش مسئولان ایرانی در تجربیات برگزیدهٔ بین المللی در زمینهٔ مبارزه با پولشویی و کمک مالی به تروریسم تأسیس شده است."



DeepMind

UNIVERSITY OF CAMBRIDGE

SMT not Science?

The COLING Paper Review **(1988)**

The validity of statistical (information theoretic) approach to MT has indeed been recognized, as the authors mention, by Weaver as early as 1949. And was universally recognized as mistaken by 1950. (cf. Hutchins, MT: Past, Present, Future, Ellis Horwood, 1986, pp. 30ff. and references therein) The crude force of computers is not science. The paper is simply beyond the scope of COLING.

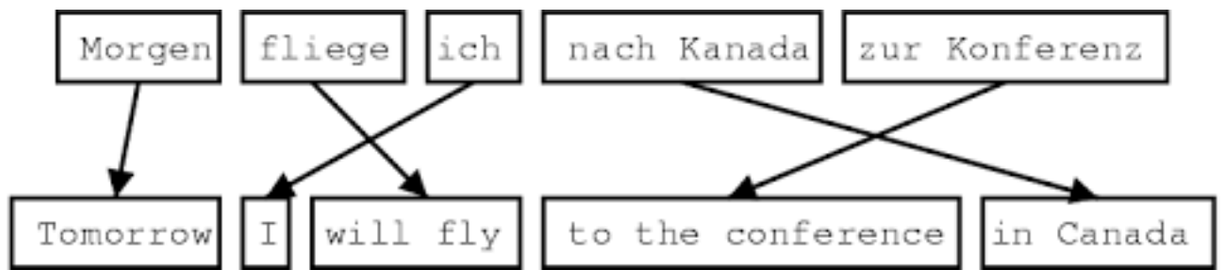
Taken from cs.jhu.edu/~post/bitext

DeepMind

UNIVERSITY OF CAMBRIDGE

Phrase-Based SMT

- 2003 onwards
- Count-based phrasal translation
- Used in e.g. the popular Moses SMT system, and previously Google translate

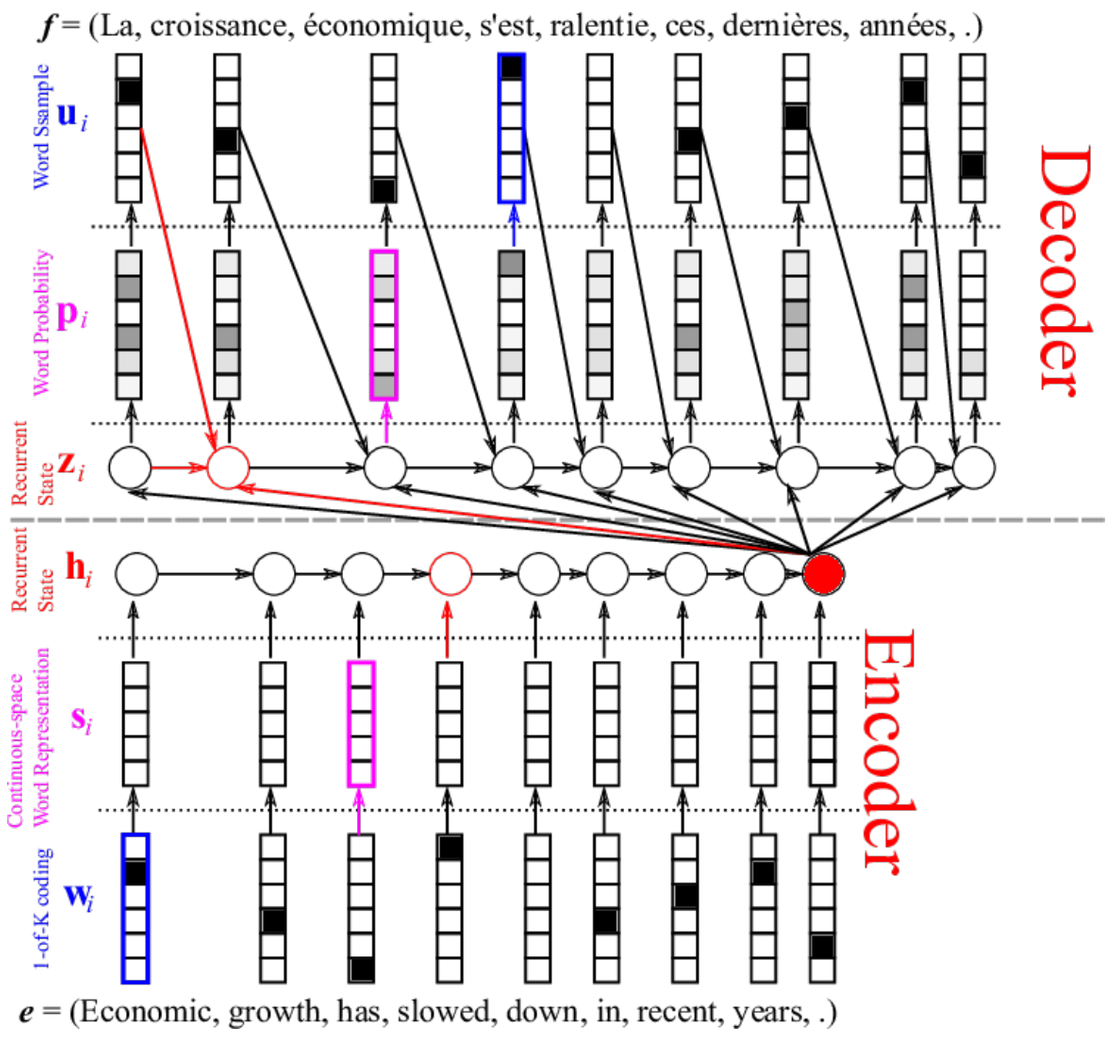


DeepMind

UNIVERSITY OF CAMBRIDGE

Neural MT

- 2013 onwards
- Based on continuous, distributed representations
- Whole system learned end-to-end using gradient descent
- Now used by Google (for many languages)

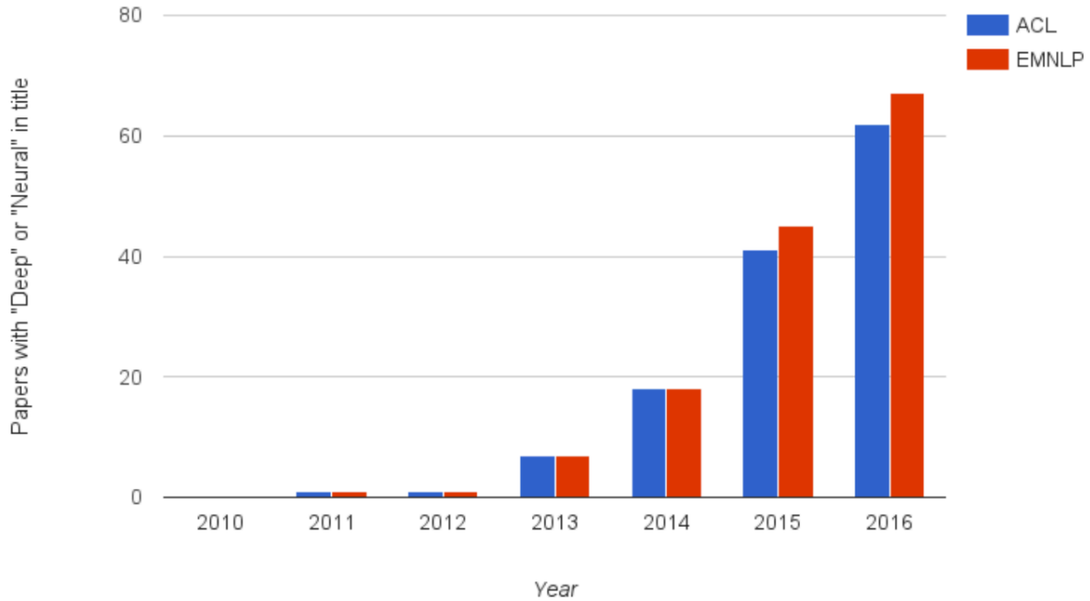


DeepMind

UNIVERSITY OF CAMBRIDGE

Is This a Revolution?

- Move from symbolic to statistical NLP (1990s) certainly was a paradigm shift
- Neural models certainly dominant in 2018
- Will neural models prove as successful for text as they have for vision and speech?



DeepMind

UNIVERSITY OF CAMBRIDGE

Remember NNs and AI/NLP have been around for decades

- Many current papers written as if NLP started in 2013
- See Jürgen Schmidhuber for a similar (highly critical, sometimes amusing) take on current NNs research

```
1 Mary moved to the bathroom.
2 John went to the hallway.
3 Where is Mary?         bathroom
4 Daniel went back to the hallway.
5 Sandra moved to the garden.
6 Where is Daniel?       hallway
7 John moved to the office.
8 Sandra journeyed to the bathroom.
9 Where is Daniel?       hallway
10 Mary moved to the hallway.
11 Daniel travelled to the office.
12 Where is Daniel?      office

FAIR bAbi task
```

DeepMind

UNIVERSITY OF CAMBRIDGE

Lectures*

1. Introduction to neural networks for NLP (Stephen Clark)
2. Feedforward neural networks (Clark)
3. Training and optimization (Clark)
4. Word embeddings (Felix Hill, DeepMind)
5. Recurrent neural networks (Hill)
6. Long short-term memory networks (Hill)
7. Convolutional neural networks (Hill)
8. Tensorflow I (Clark)
9. Conditional language models (Chris Dyer, DeepMind and CMU)
10. Conditional language models with attention (Dyer)
11. Machine comprehension (Ed Grefenstette, DeepMind)
12. Tensorflow II (Clark)
13. Sentence representations and inference (Grefenstette)
14. Image captioning (Clark)
15. Grounded language learning (Hill)
16. Language in Labyrinth (Hill)

***subject to change**

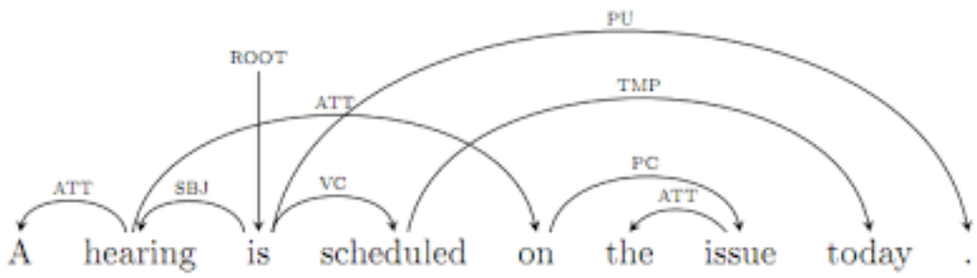
DeepMind

UNIVERSITY OF CAMBRIDGE

What this course is not about

- Not about building tools for intermediate representations (parsers, pos taggers)

- More focus on “end-to-end” task-based training



DeepMind

UNIVERSITY OF CAMBRIDGE

AI-hard Tasks/Applications

Language Understanding

CNN article:

Document	The BBC producer allegedly struck by Jeremy Clarkson will not press charges against the "Top Gear" host, his lawyer said Friday. Clarkson, who hosted one of the most-watched television shows in the world, was dropped by the BBC Wednesday after an internal investigation by the British broadcaster found he had subjected producer Oisin Tymon "to an unprovoked physical and verbal attack." ...
Query	Who does the article say will not press charges against Jeremy Clarkson?
Answer	Oisin Tymon

thanks to Phil Blunsom for slide

DeepMind

UNIVERSITY OF CAMBRIDGE

AI-hard Tasks/Applications

Image Understanding

What is the man holding?
Does it appear to be raining?
Does this man have 20/20 vision?

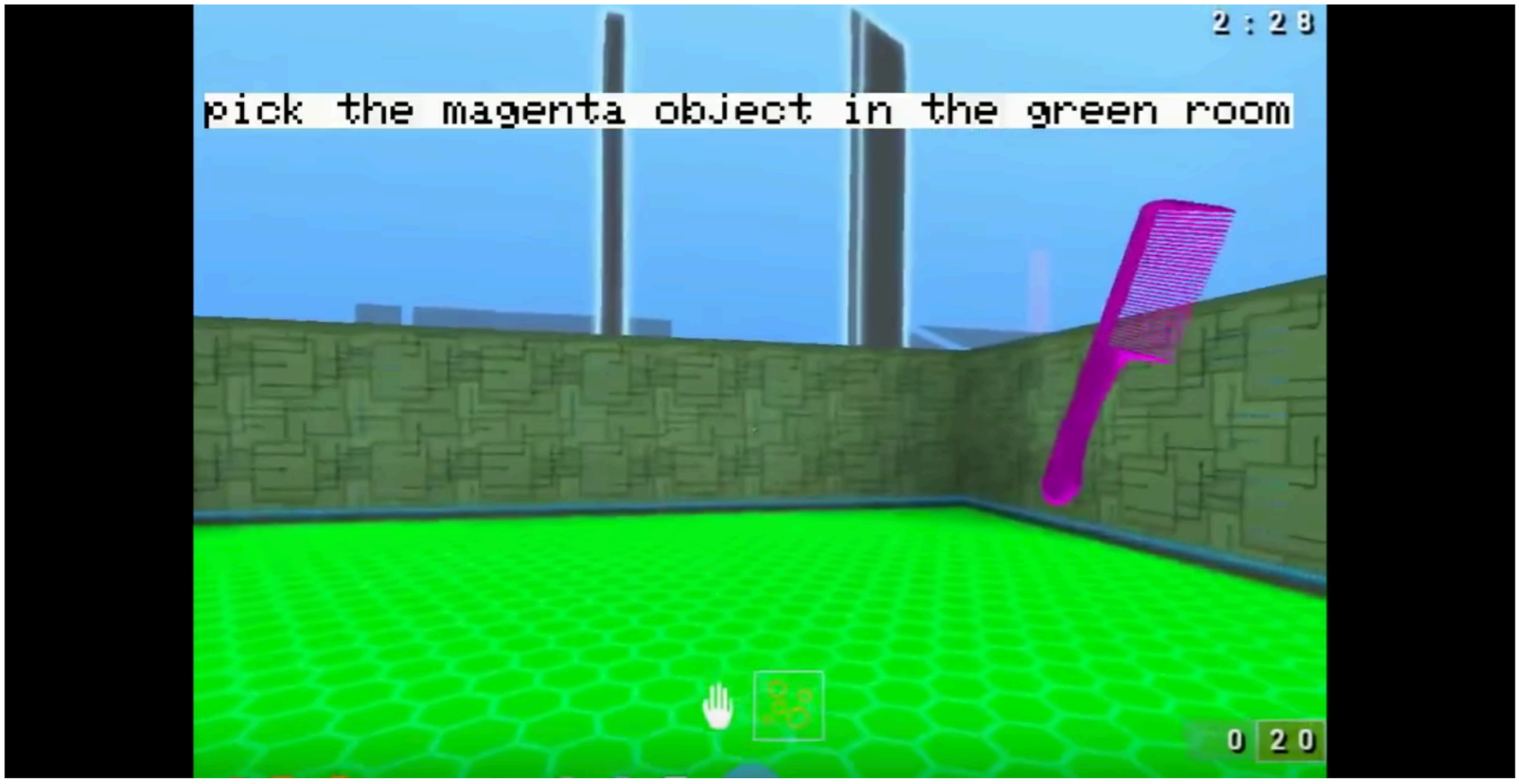
thanks to Phil Blunsom for slide

DeepMind

UNIVERSITY OF CAMBRIDGE

AI-hard Tasks/Applications

Grounded Language Learning



https://www.youtube.com/watch?v=wJjdu1bPJ04&feature=youtu.be

DeepMind

UNIVERSITY OF CAMBRIDGE

Practical

- Experimenting with a dictionary definition model using Tensorflow

- Microsoft Azure cloud resources available

- Two 2-hour sessions, weeks 4 and 6 on Tuesday afternoons

DeepMind

UNIVERSITY OF CAMBRIDGE

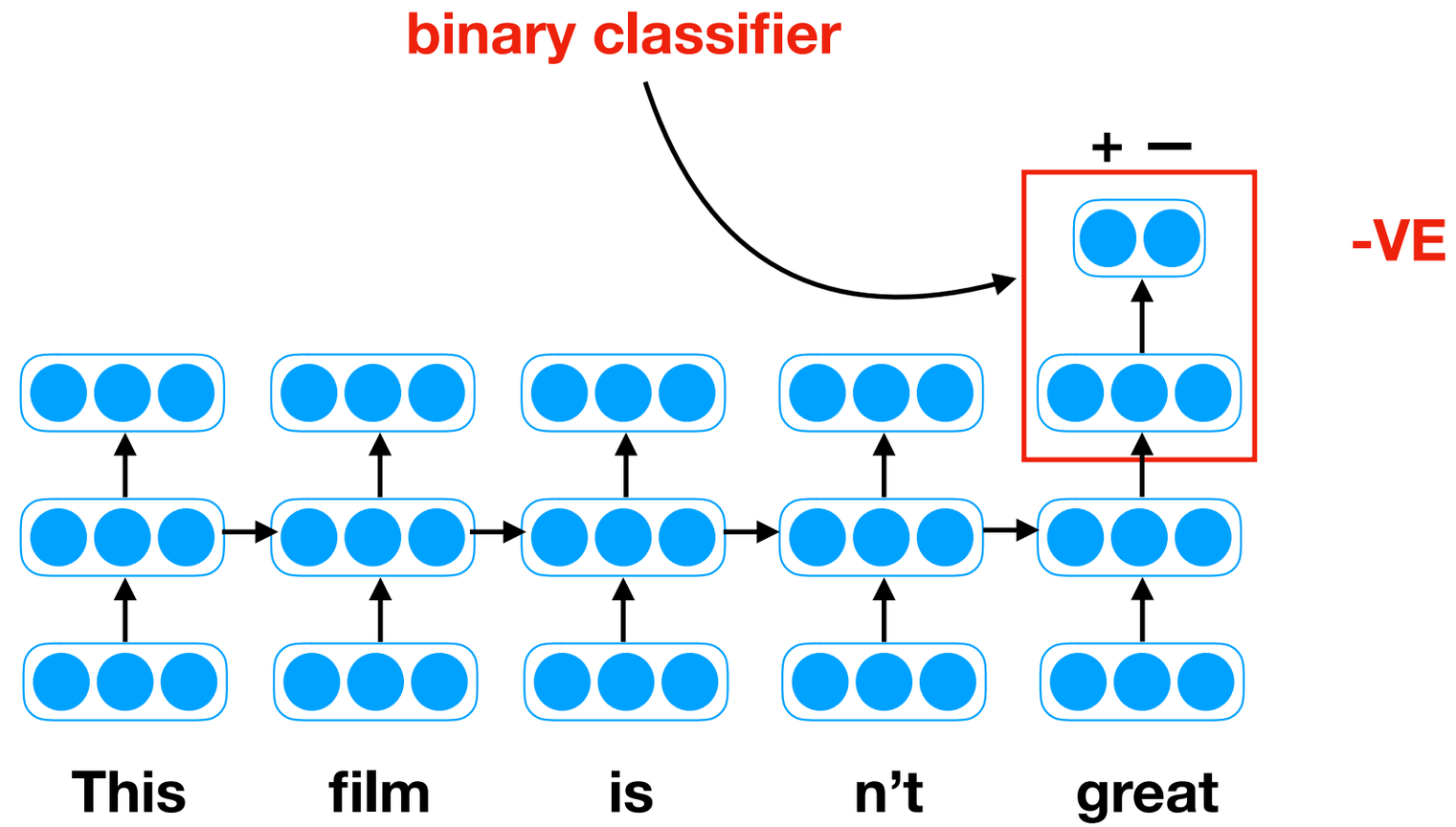
Assessment

- One practical report (40%)
- One take-home exam (60%)
- See course web pages for details on when these are due

DeepMind

UNIVERSITY OF CAMBRIDGE

Components of an End-to-End (Sentiment Analysis) System

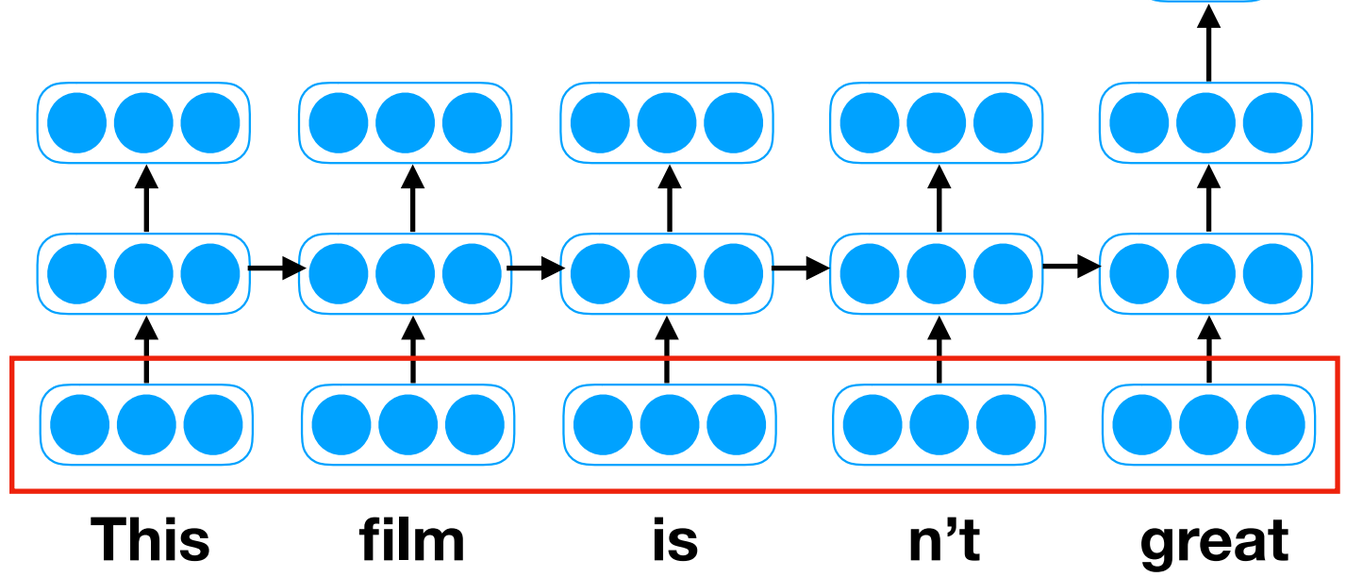


DeepMind

UNIVERSITY OF CAMBRIDGE

Word Embeddings

- Random initialization, learn as part of task objective
- External initialization (eg Word2Vec), update as part of task objective
- External initialization, keep fixed



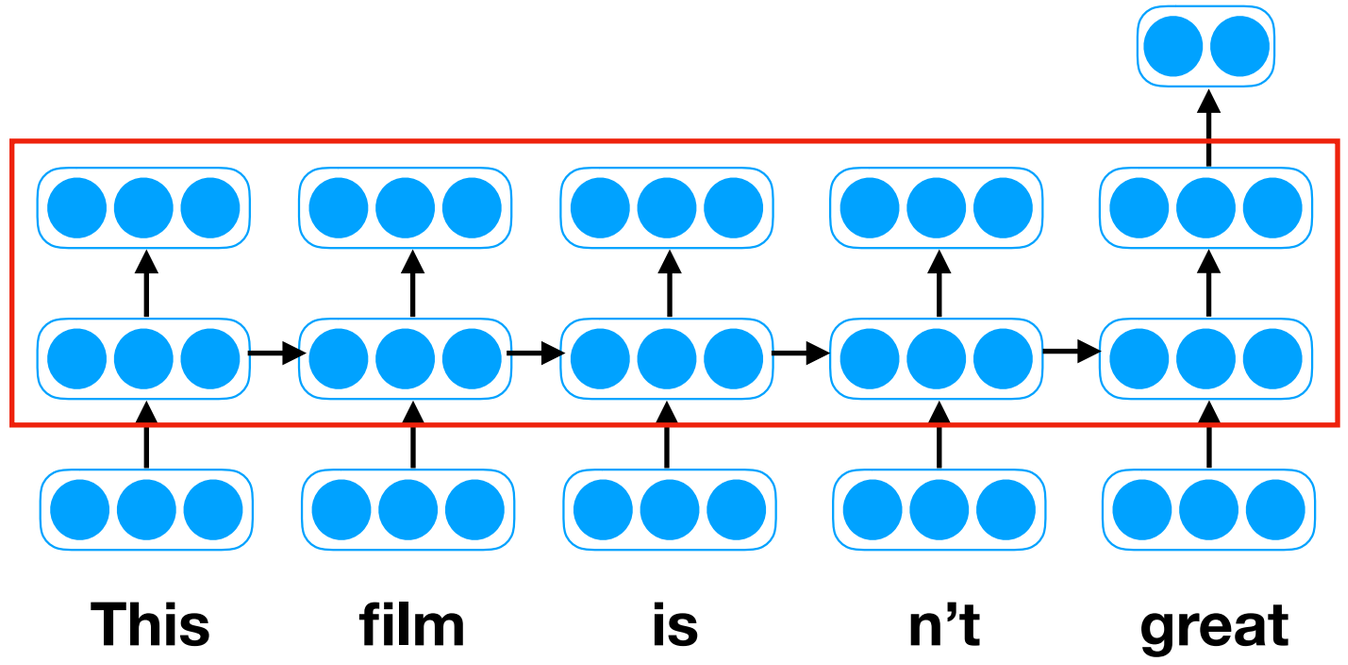
DeepMind

UNIVERSITY OF CAMBRIDGE

Sentence Embeddings

- Recurrent neural network (RNN, LSTM, Tree RNN) combines the word vectors

- Could use a convolutional neural network (CNN), or a combination of RNN, CNN

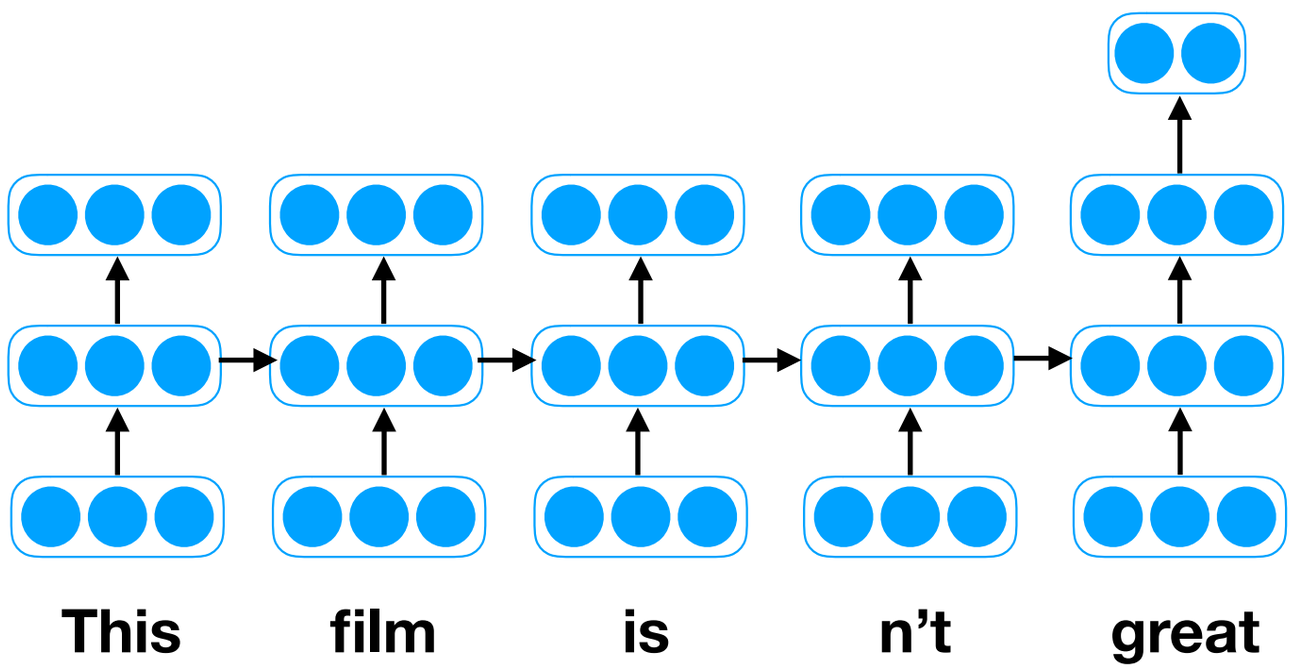


DeepMind

UNIVERSITY OF CAMBRIDGE

Why on Earth Does it Work?

- Huge number parameters (so need lots of data)
- Distributed representations share statistical strength
- Optimization surface is highly non-convex, system is highly non-linear (but SGD works surprisingly well)



DeepMind

UNIVERSITY OF CAMBRIDGE

Main Readings

- *Deep Learning,* Goodfellow, Bengio and Courville
  www.deeplearningbook.org

- *A Primer on Neural Network Models for Natural Language Processing,* Yoav Goldberg

DeepMind

UNIVERSITY OF CAMBRIDGE