UE903 EC1: Application to Text

Natural Language Generation (NLG)

Introduction

Claire Gardent

Claire Gardent CNRS / LORIA

Today

Course Logistics

What is Text Production ?

Pre-Neural Approaches to Text Production

Logistics

UE903

Course Logistics

Setting Up the Scene

- Two Lectures

Studying the various Approaches

- Presentations and Quizzes

Final Exam

- 1 hour: NLG (C. Gardent)
- 1 hour: Sentiment Analysis, etc. (D. Langlois)

Contact and Documentation

- Email: claire.gardent@loria.fr
- Website: https://members.loria.fr/CGardent/teaching/ue903/README.html

UE903

Course Logistics

Schedule and syllabus

Event type	Date	Time	Description	Course material
Session 1	10/09	10-12pm	**Introduction to Text Production**	[slides]
			Pre-Neural Approaches to Text Production	[slides]
Session 2	17/09	10-12am	**Neural Approaches to Text Production**	[slides]
Session 3	24/09	2-5pm	**Presentations + Quiz**	
Session 4	26/09	10-12pm	**Presentations + Quiz**	
Session 5	01/10	10-12pm	**Presentations + Quiz**	
Session 6	03/10	10-12pm	**Presentations + Quiz**	
Exam	06/02	2-4pm		

UE903

Course Logistics

Presentations + Quiz

- Presentations
  - pre-assigned
  - 10 minutes + 5 mn Q&A
  - Based on scientific paper
  - Slides sent to the class at the latest at 12pm on the day before the presentation
  - graded
- Quiz on the presented scientific papers
  - 15 minutes per paper (usually half an hour for 2 papers)
  - slides and paper are available
  - graded

UE903

Course Logistics

Presentations + Quiz

- Reading List
- Each student presents one paper (see List)
- All students read all papers
- The slides for the presentation must be sent to the lecturer AND to the students at the latest at 12pm before the presentation day
- After the presentations, a written quiz must be answered by each student . Usually the quiz will bear on two (related) presentations. So roughly, there will be half an hour of talks and Q/A followed by half an hour of written quiz.
- Presentation: 10', Q&A: 5', Quiz: 15'

UE903

Course Logistics

Grading

- 50% Written Exam
- 10% Presentation
- 40% Quiz scores

Questions ?

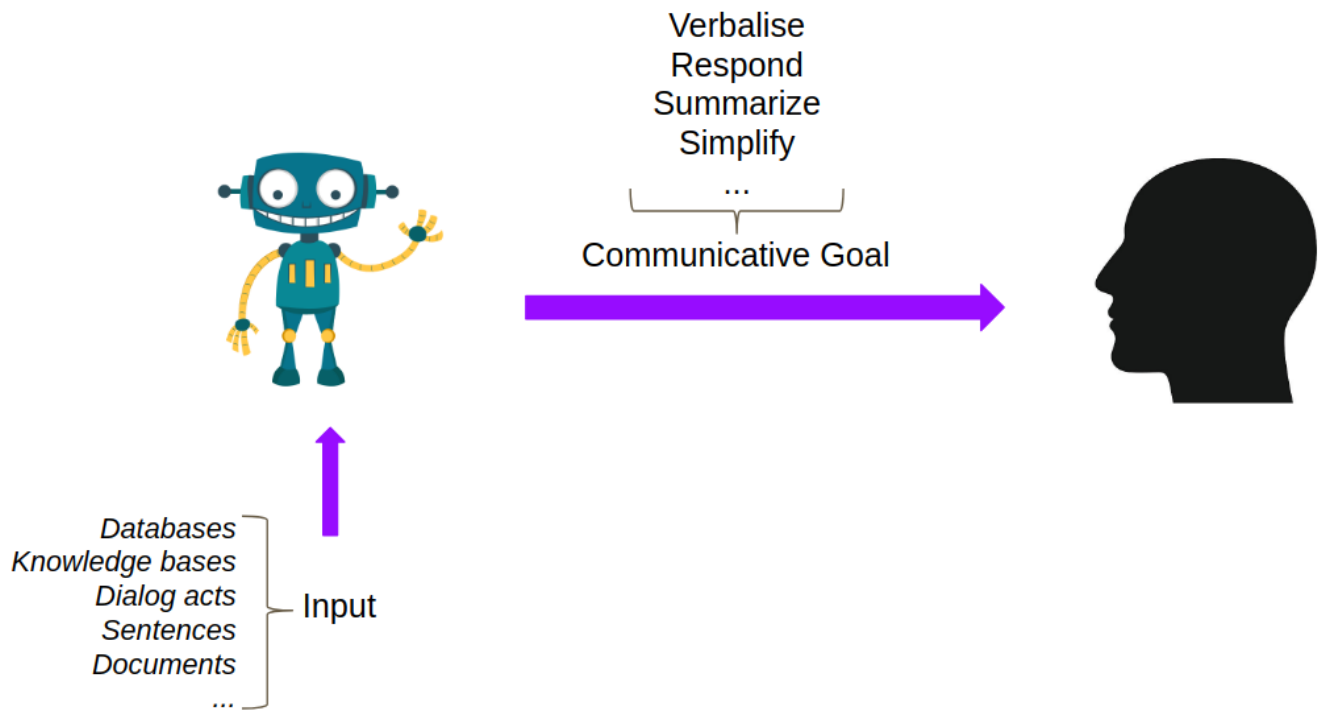
UE903

Course Logistics

Introduction

What is Text Production ?

Input and goals define different types of text production tasks



Three Main Types of Input

Meaning Representations

- Abstract Meaning Representations
- Dependency Trees (Deep and Shallow)
- Discourse Representation Structures
- Logical Formulae (First Order Logic, Description Logic)

Data

- Knowldege Bases
- Data Bases
- Numerical data from signal processing

Text

- Short, Long
- Multiple or single document
- Dialog or Discourse

Various Types of Communicative Goals

Describing, Verbalising

- a KB fragment
- an entity in a DB
- an image, a video

Summarising

- A text
- Several texts
- The content of a KB

Simplifying

- For children, foreigners, disabled people

Paraphrasing, Reformulating

- Expert/non expert
- Varied chatbot output

UE903

Example 1: Verbalising Knowledge Base Fragments

The WebNLG Challenge 2017

(John_E_Blaha *birthDate* 1942_08_26) (John_E_Blaha *birthPlace* San_Antonio) (John_E_Blaha *occupation* Fighter_pilot)

"John E Blaha, born in San Antonio on 1942-08-26, worked as a fighter pilot."

UE903

Course Logistics

Introduction

Example Applications

Example 2: Generating from Data Bases

(a)

Flight		Search		Day	
From	To	Type	What	Day	Dep/Ar
phoenix	new_york	query	flight	sunday	departure

List flights from phoenix to new york on sunday

(b)

Temperature			
Time	Min	Mean	Max
06:00-21:00	9	15	21

Cloud Sky Cover	
Time	Percent (%)
06:00-09:00	25-50
09:00-12:00	50-75

Wind Speed			
Time	Min	Mean	Max
06:00-21:00	15	20	30

Wind Direction	
Time	Mode
06:00-21:00	S

Cloudy, with a low around 10. South wind around 20 mph.

(c)

Pass		Bad Pass		Turn Over	
From	To	From	To	From	To
pink3	pink7	pink7	purple3	pink7	purple3

pink3 passes the ball to pink7

Angeli, Liang and Klein 2010 ; Konstas and Lapata 2012

UE903

Course Logistics

Introduction

Example Applications

Example 3: Captioning Images and Videos

Vinyals et al., 2015

UE903

Course Logistics

Introduction

Example Applications

Example 4: Summarising

REVIEW

doi:10.1038/nature14539

Deep learning

Yann LeCun[1,2], Yoshua Bengio[3] & Geoffrey Hinton[4,5]

Deep learning allows computational models that are composed of multiple processing layers to learn representations of data with multiple levels of abstraction. These methods have dramatically improved the state-of-the-art in speech recognition, visual object recognition, object detection and many other domains such as drug discovery and genomics. Deep learning discovers intricate structure in large data sets by using the backpropagation algorithm to indicate how a machine should change its internal parameters that are used to compute the representation in each layer from the representation in the previous layer. Deep convolutional nets have brought about breakthroughs in processing images, video, speech and audio, whereas recurrent nets have shone light on sequential data such as text and speech.

Machine-learning technology powers many aspects of modern society: from web searches to content filtering on social networks to recommendations on e-commerce websites, and it is increasingly present in consumer products such as cameras and smartphones. Machine-learning systems are used to identify objects in images, transcribe speech into text, match news items, posts or products with users' interests, and select relevant results of search. Increasingly, these applications make use of a class of techniques called deep learning.

Conventional machine-learning techniques were limited in their ability to process natural data in their raw form. For decades, constructing a pattern-recognition or machine-learning system required careful engineering and considerable domain expertise to design a feature extractor that transformed the raw data (such as the pixel values of an image) into a suitable internal representation or feature vector from which the learning subsystem, often a classifier, could detect or classify patterns in the input.

intricate structures in high-dimensional data and is therefore applicable to many domains of science, business and government. In addition to beating records in image recognition[1–4] and speech recognition[5–7], it has beaten other machine-learning techniques at predicting the activity of potential drug molecules[8], analysing particle accelerator data[9,10], reconstructing brain circuits[11], and predicting the effects of mutations in non-coding DNA on gene expression and disease[12,13]. Perhaps more surprisingly, deep learning has produced extremely promising results for various tasks in natural language understanding[14], particularly topic classification, sentiment analysis, question answering[15] and language translation[16,17].

We think that deep learning will have many more successes in the near future because it requires very little engineering by hand, so it can easily take advantage of increases in the amount of available computation and data. New learning algorithms and architectures that are currently being developed for deep neural networks will only accelerate this progress.

Grusky et al., 2018

UE903

Course Logistics

Introduction

Example Applications

Example 5: Generating Headlines

Mail Online — Science & Tech

Home | News | U.S. | Sport | TV&Showbiz | Australia | Femail | Health | Science | Money | Video | Travel | Fashion Finder

Latest Headlines | Science | Pictures | Discounts — Login

ADVERTISEMENT

Did Cambridge Analytica get YOUR data? Facebook will tell 87 million affected users TODAY if their information was shared

- Starting today all 2.2 billion Facebook users will receive a notice on their feeds
- 'Protecting Your Information,' notice will contain link to see what apps they use
- This will also show what information they have shared with those apps, firm says
- Users whose data may have been shared with Cambridge Analytica will be told
- Facebook says most of the 87 million affected users are in the United States
- In an interview Sunday, Cambridge Analytica whistleblower Christopher Wylie said the number could actually be larger than 87 million

By ASSOCIATED PRESS
PUBLISHED: 01:14, 9 April 2018 | **UPDATED:** 08:58, 9 April 2018

Share 930 shares — 158 View comments

Today Facebook will tell the 87 million users who may have had their information shared with Cambridge Analytica.

Starting Monday all 2.2 billion **Facebook** users will receive a notice on their newsfeeds, titled 'Protecting Your Information,' with a link to see what apps they use and what information they have shared with those apps.

If they want, they can shut off apps individually or turn off third-party access to their apps completely.

Site Web Enter your search Search

ADVERTISEMENT

Like +1

Chopra et al., 2016

UE903

Example 6: Summarising Multiple Documents

India backs down on proposed "fake news" legislation after an outcry

- On Tuesday, the Indian government walked back a new rule that would have punished publishers of so-called "fake news," after many questioned what exactly would fall into that category. [CNN / Sugam Pokharel and Joshua Berlinger]
- On Monday, the government announced that journalists who were found to have written "fake news" would lose their official accreditation, in some cases permanently. But the proposal faced such swift and strong backlash that by Tuesday, the government had changed its tune. [NYT / Kai Schultz and Suhasini Raj]
- Many in the Indian news media saw the new rules as an attack on the press, noting organizations like the Press Council of India and the News Broadcasters Association already exist to ensure press accountability. [Times of India]
- Indian journalists also pointed out that the amendment was released mere months before campaigning was set to begin for national elections in 2019, and that Prime Minister Narendra Modi's party has a history of attacking members of the press who publish criticism of their leadership. [Times of India]

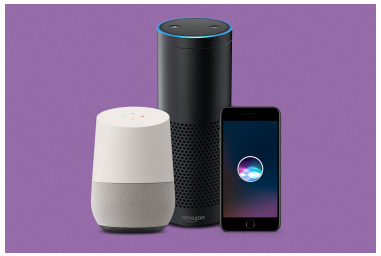
UE903

Course Logistics

Introduction

Example Applications

Example 7: Conversational Agents



Li et al., 2016a; Li et al., 2016b

A: Where are you going ?
B: I'm going to the police station.
A: I'll come with you
B: No, no, non, you're not going anywhere
A: Why?
B: I need you to stay here
A: I don't know what you are talking about.

UE903

NLG

- Many different inputs
  Data, Meaning Representations, Text
- Many different communicative goals
  Verbalise, summarise, compress, simplify, respond, compare ….

Some Terminology

- D2T Generation: Generating from Data
- MR2T Generation: Generating from Meaning Representations
- T2T Generation: Generating from Text
  Summarisation, Compression, Paraphrasing, Simplification