WARWICK

CS918: LECTURE 17

Text Summarisation

Arkaitz Zubiaga, 28th November, 2018

LECTURE 18: CONTENTS

- What is Text Summarisation? Applications.
- Text Summarisation Tasks:
  - Single-Document Summarisation.
  - Multi-Document Summarisation.
- Evaluation of Text Summarisation:
  - ROUGE.

WHAT IS TEXT SUMMARISATION?

WARWICK

- **Text summarisation:** To extract the gist, the essence, of a text and present it in a **shorter form with as little loss as possible** with respect to mediated information.

- Longer, source text → shorter, target text.

- Automatic text summarisation is an old task in NLP, with the earliest systems dating back to the 1960's.

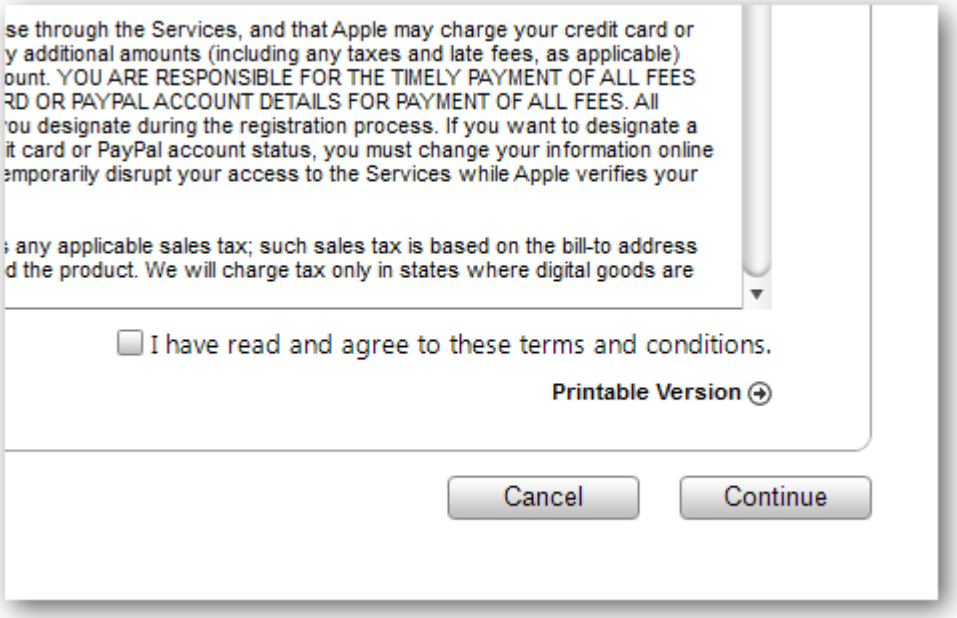
TEXT SUMMARISATION: APPLICATIONS

WARWICK

- **Summarise news**. Useful both for readers and for journalists, to identify key facts.
- **Summarise reviews** on products.
- **Summarise research papers**, too much to read!
- **Text simplification**, e.g. for non-experts.
- **For search engines**, instead of snippets, show summaries.
- **Question answering**.

TEXT SUMMARISATION: APPLICATIONS

- **TL;DR**



METHODS OF TEXT SUMMARISATION

- Two main methods for text summarisation:
  - **Extractive summarisation:**
    - Identify most important segments (e.g. sentences, paragraphs) in source text.
    - Those segments will compose the target text.
  - **Abstractive summarisation:**
    - Identify the key information of the source text.
    - Using that information, write new sentences that will form the target text.

EXTRACTIVE VS ABSTRACTIVE SUMMARIES

- Source text: This text is too long. I want a summary of this text. The text will then be shorter. More people will read the shorter text.

- Extractive summary: I want a summary of this text. More people will read the shorter text.

- Abstractive summary: More people will read a shortened version of this long text.

EXTRACTIVE VS ABSTRACTIVE SUMMARIES

WARWICK

- Ideally, **we'd prefer an abstractive summary, but it's much more difficult** to create.
  - With **extractive** summaries, the challenge lies in **making the resulting text cohesive**.
  - With **abstractive** summaries, the main challenge lies in **generating sentences** (NLG, natural language generation).

MORE TYPES OF SUMMARIES

- **Input type:**
  The input can be one document (**single-document**), or multiple documents (**multi-document**). With multiple documents, need to avoid repetition in the summary.

- **Purpose of summary:**
  Indicative (i.e. provide cursory understanding of a topic) vs **informative** (i.e. detailed breakdown of text).

- **Focus of summary:**
  Generic vs. **Query-based** (i.e. key information of web page that relates to a user query, useful for question answering) vs. **Domain-specific** (e.g. medical text)

VERY SIMPLE BASELINE

- Take the first sentence as the summary.

United Kingdom

From Wikipedia, the free encyclopedia
(Redirected from UK)

This article is about the country. It is not to be confused with Great Britain, its largest island whose name is also loosely applied to the whole country.
"UK" redirects here. For other uses of "UK", see UK (disambiguation). For other uses of "United Kingdom", see United Kingdom (disambiguation).

The **United Kingdom of Great Britain and Northern Ireland**, commonly known as the **United Kingdom** (**UK**) or **Britain**, is a sovereign country in western Europe. Lying off the north-western coast of the European mainland, the UK includes the island of Great Britain, the north-eastern part of the island of Ireland and many smaller islands.[10] Northern Ireland is the only part of the United Kingdom that shares a land border with another sovereign state—the Republic of Ireland.[note 9] Apart from this land border, the UK is surrounded by the Atlantic Ocean, with the North Sea to its east, the English Channel to its south and the Celtic Sea to its south-south-west, giving it the 12th-longest coastline in the world. The Irish Sea lies between Great Britain and Ireland. With an area of 242,500 square kilometres (93,600 sq mi), the UK is the 78th-largest sovereign state in the world. It is also the 21st-most populous country, with an estimated 65.5 million inhabitants in 2016.

FOCUS ON EXTRACTIVE SUMMARISATION

- In this lecture we will be focusing on extractive summarisation:
  - Single-document summarisation.
  - Multi-document summarisation.

WARWICK

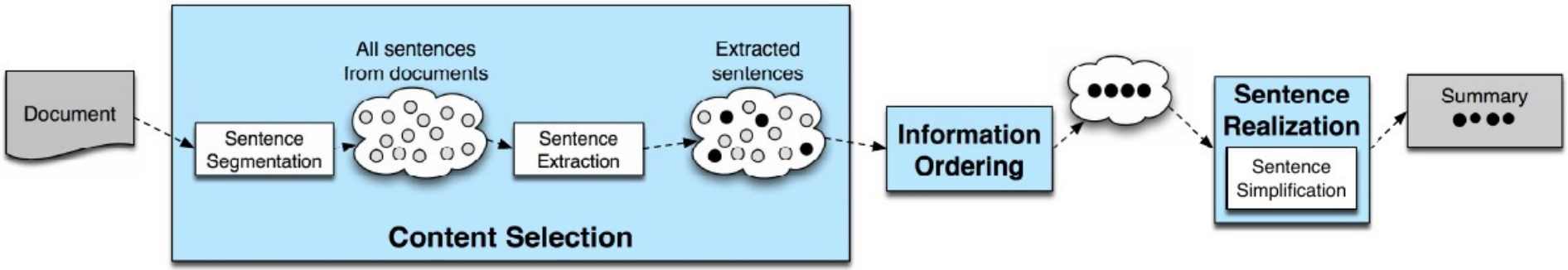
EXTRACTIVE, SINGLE-DOCUMENT SUMMARISATION

SINGLE-DOCUMENT SUMMARISATION

WARWICK

- Let's assume the segmentation we perform on the source text is into sentences. We then perform three steps:
  1. **Content Selection:** Choose sentences to extract.
  2. **Information Ordering:** Decide how to order the sentences.
  3. **Sentence Realisation:** Clean up sentences, remove non-essential phrases from sentences, fuse several sentences into one, fix problems in coherence.

SINGLE-DOCUMENT SUMMARISATION: ARCHITECTURE



STEP 1: CONTENT SELECTION

- Often treated as a classification task: binary classification as “extractworthy” vs “non-extractworthy”.

- This classification task can be unsupervised or supervised.

STEP 1: UNSUPERVISED CONTENT SELECTION

- For each word in the source text, determine its **saliency**.
  - We can determine the saliency of w_i using TF-IDF:

$$weight(w_i) = \mathrm{tf}_{i,j} \times \mathrm{idf}_i$$

- Each of the k sentences in the input is then assigned a **centrality score**, i.e. average cosine with all other sentences:

$$centrality(x) = \frac{1}{K} \sum_{y} tf\text{-}idf\text{-}cosine(x, y)$$

- We can then choose to **include the most "central" sentences** in the summary.

STEP 1: SUPERVISED CONTENT SELECTION

WARWICK

- Need **documents and human-created summaries** as input.
  - In the document, we'll label sentences in the summary as 1, all other sentences as 0.

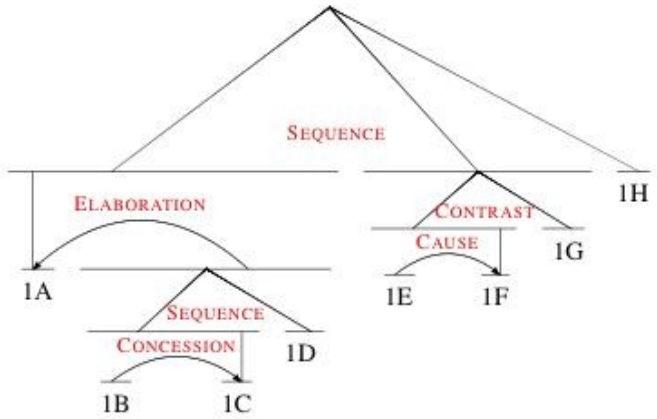
- We can **train a Naive Bayes or Maximum Entropy classifier**, when applied to unseen documents, they'll output the probability that each sentence is extractworthy.
  - We can use features like:
    - Position (e.g. position of sentence, is it first sentence?).
    - Length of sentence.
    - Informativeness of its words (e.g. tf-idf).

STEP 2: ORDERING OF SENTENCES

- **Rhetorical Structure Theory (RST)** can be used to perform discourse analysis.
  - It can help determine discursive roles of sentences, such as nucleus, satellites, background, justification, evidence, cause, concession, etc.
  - Discursive roles of sentences can help:
    - decide the order.
    - make sentences cohesive.

STEP 2: ORDERING OF SENTENCES

- **Rhetorical Structure Theory (RST)**



[Yesterday, the delegates chose their new representative.]^{1A} [Even though Smith received only 24 votes,]^{1B} [he accepted the election with a short speech.]^{1C} [Then the assembly applauded for three minutes.]^{1D} [Due to the upcoming caucus meeting,]^{1E} [the subsequent discussion was very short.]^{1F} [Nonetheless the most pressing questions could be resolved.]^{1G} [The meeting was closed at 7pm.]^{1H}

STEP 2: ORDERING OF SENTENCES

- **Alternative:** keep the original order!
  - After all we've used a single document as input.
  - We can assume that the content was well structured and ordered in the original document.

STEP 3: SENTENCE REALISATION

- Simplification of long sentences:
  - Original sentence: ~~When in arrives sometime new year in new TV sets~~, the V-chip will give parents a ~~new and potentially revolutionary~~ device to block out programs they don't want their children to see.
  - Simplified sentence: The V-chip will give parents a device to block out programs they don't want their children to see.

STEP 3: SENTENCE REALISATION

- Simplest approaches are rule-based to remove:
  - **Appositives:** Rajam, ~~an artist living in Philadelphia~~, found the inspiration in the back of city magazines.
  - **Attribution clauses:** Rebels agreed to talk with government officials, ~~international observers said Tuesday~~.
  - **PPs with no named entities:** The commercial fishing restrictions in Washington will not be lifted unless the salmon population increases ~~to a sustainable number~~.
  - **Initial adverbials:** for example, on the other hand, as a matter of fact, at this point.
- Alternatively, machine learning can be used over manually labelled sentences.

EXTRACTIVE, MULTI-DOCUMENT SUMMARISATION

MULTI-DOCUMENT SUMMARISATION

WARWICK

- How is it useful?
  - Summarise multiple **news articles** on the same story.
  - **Question answering** when answers need combining content from multiple web pages.
  - Summarise what's being said across **social media** posts.

MULTI-DOCUMENT SUMMARISATION

WARWICK

- It’s far from solved, but current systems can be **useful for information-finding tasks**.

- We need a set of **related documents as input**.

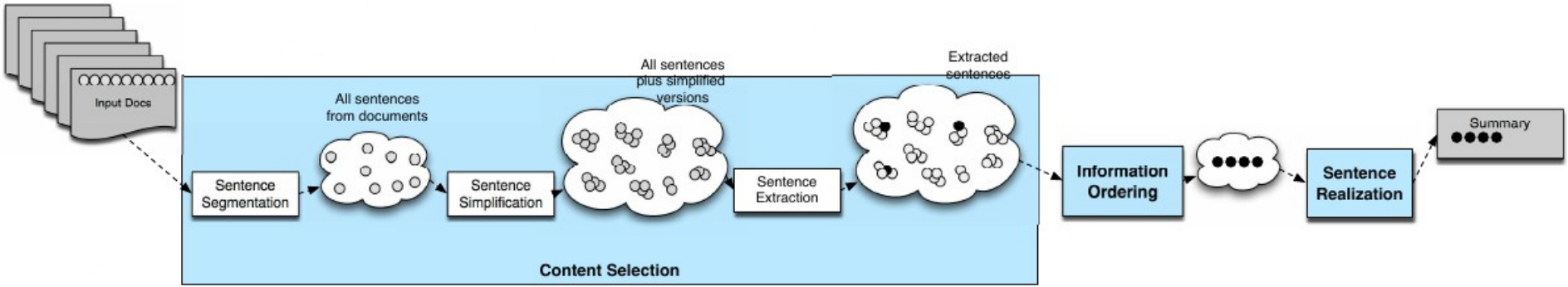
MULTI-DOCUMENT SUMMARISATION

WARWICK

- Three steps, as in single-document summarisation:
  - Content selection.
  - Information ordering.
  - Sentence realisation.

MULTI-DOCUMENT SUMMARISATION

- Content selection is more complex.



STEP 1: CONTENT SELECTION

- Issue: **dearth of labelled corpora** to train supervised classifier.
  - Generally performed as an **unsupervised task**.
  - Main challenge in multi-document is **getting rid of redundancy** across documents.

STEP 1: CONTENT SELECTION

- **Maximal Marginal Relevance (MMR)** for redundancy removal.
  - **Penalisation factor** based on similarity between sentence s and the set of sentences already selected for the summary.

MMR penalisation(s) = $\lambda max_{s_i \in Summary} \mathrm{Sim}(s, s_i)$

where λ is a predefined parameter, and sim() is some similarity scoring function (e.g. cosine similarity)

STEP 2: INFORMATION ORDERING

- Keeping the **original order** (as in single-document summaries) is **not an option here**.
- We can do **chronological ordering**, using document timestamps.
  - Good baseline, but summaries tend to **lack cohesion**.
  - Besides, determine **lexical cohesion** between sentences.
    - This can be done by putting pairs of sentences with **low TF-IDF cosine distance next to each other**.
- Other approaches based on parsing and dependencies have been proposed, but this is still an **open research problem**.

STEP 3: SENTENCE REALISATION

WARWICK

- We can use the same techniques as for single-document.

- Additionally, we need **coherence in named entity mentions**.
  - For instance, if "**UK Prime Minister Theresa May**" and "**May**" are both mentioned in a summary.

  - We need to make sure that the **full name appears first**, the last name only appears later.

WARWICK

EVALUATION OF TEXT SUMMARISATION

DATASETS

- There aren’t many datasets for text summarisation.

- The Document Understanding Conference (DUC) used to run a shared task on text summarisation until 2007, and data was released: http://duc.nist.gov/data.html

DATASETS

- Use other corpora as a proxy for summarisation, e.g.:
  - **Scientific papers** → assume the abstract is the summary you want to aim for.
  - **News articles** → assume the lead paragraph is a good summary for evaluation.
    e.g. http://homepages.inf.ed.ac.uk/mlap/resources/cnnhlights/

EVALUATION OF SUMMARIES

- Evaluation is a challenge!
  - Contrary to other tasks like classification, there are **many possible and valid summaries**.

- Evaluation can be **extrinsic** (task-based) **or intrinsic** (task-independendent).

EXTRINSIC EVALUATION

- Evaluation can be done involving people.
  - Prepare a **set of questions** that can be answered by reading the entire document(s).
  - Have different people **read the summaries**.
  - How many of the questions are they **able to answer?**

INTRINSIC EVALUATION

- We need a corpus with **human-generated, reference summaries**.
  - Ideally multiple reference summaries for each document for generalisation.

- Best-known evaluation metric is **ROUGE:** Recall-Oriented Understudy for Gisting Evaluation.

  It is based on Bleu, an evaluation metric used for Machine Translation, which computes the overlap of (all of) unigrams, bigrams, trigrams and quadrigrams between reference and automatic translations.

ROUGE

- The **length of n-grams is fixed**, ROUGE-1 for unigrams, ROUGE-2 for bigrams, and so on.

$$ROUGE2 = \frac{\sum_{S \in \{ReferenceSummaries\}} \sum_{bigram \in S} \text{Count}_{\text{match}}(bigram)}{\sum_{S \in \{ReferenceSummaries\}} \sum_{bigram \in S} \text{Count}(bigram)}$$

- ROUGE-2: out of all bigrams in the reference summaries, how many does the automatic summary have?

ROUGE: EXAMPLE

- Ref 1: Water spinach is a green leafy vegetable grown in the tropics. [3/10]
- Ref 2: Water spinach is a semi-aquatic tropical plant grown as a vegetable. [3/9]
- Ref 3: Water spinach is a commonly eaten leaf vegetable of Asia. [6/9]

- System answer: Water spinach is a leaf vegetable commonly eaten in tropical areas of Asia.

ROUGE2 = (3 + 3 + 6) / (10 + 9 + 9) = 12 / 28 = .43

AN ALTERNATIVE: THE PYRAMID METHOD

- In reference summaries, we get annotations for units of meaning called **Summary Content Units** (SCU), i.e. sequences of words that we want in the summaries, e.g.

> The industrial espionage case involving GM and VW began with the hiring of Jose Ignacio Lopez, an employee of GM subsidiary Adam Opel, by VW as a production director.

- Then we can measure the overlap of SCUs, which we can complement with n-grams of predefined n for the rest.

LIMITATIONS OF ROUGE (AND INTRINSIC METHODS)

WARWICK

- We **only look at n-gram overlap**, we don't measure cohesion, coherence, grammaticality, etc.

- Ideally we'd have **people look at the summaries** too:
  - Are they grammatically correct?
  - Does the summary read well (cohesion, coherence)?
  - Can you answer some questions?

SUMMARISATION: SUMMARY

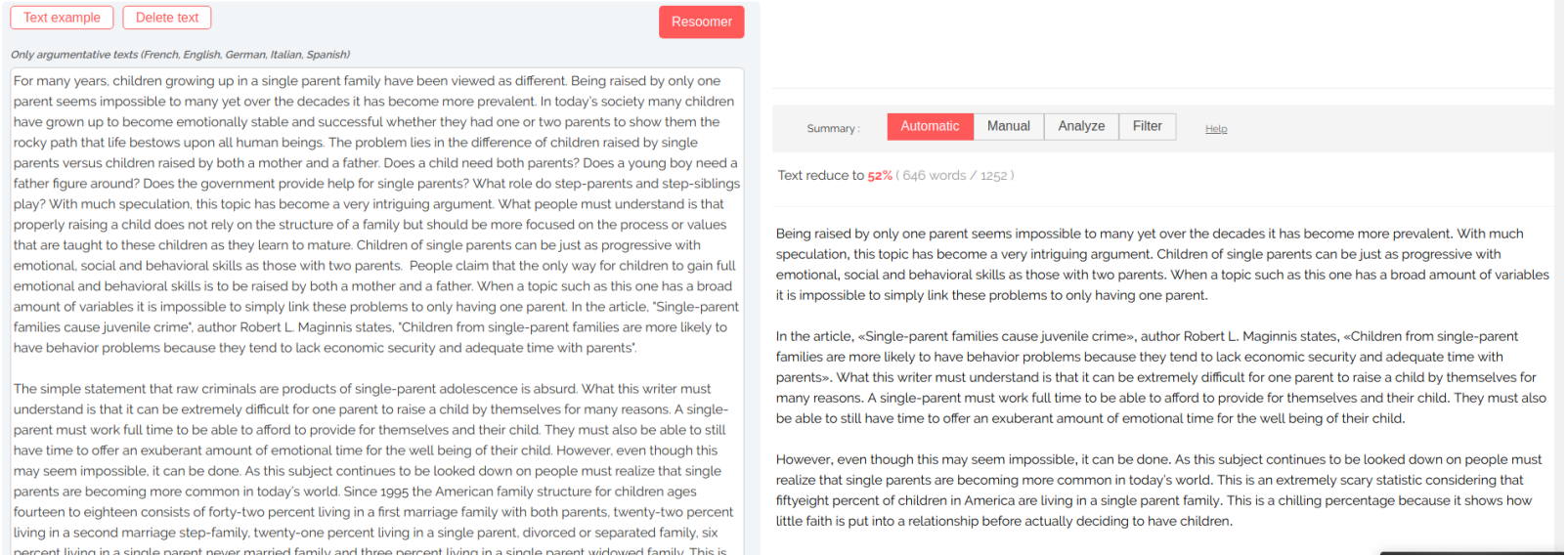
- **Types of summaries:**
  - Single- vs multi-document.
  - Extractive vs abstractive.
- **Pipeline:**
  - Content selection.
  - Information ordering.
  - Sentence realisation.
- **Current methods mostly focus on supervised content selection.**
  - A long way to go in deeper linguistic understanding of content for summarisation.
  - Likewise for more advanced evaluation methods that consider syntax, grammar, cohesion.

RESOURCES

- gensim.summarization:
  https://rare-technologies.com/text-summarization-with-gensim/
- PyTeaser:
  https://github.com/xiaoxu193/PyTeaser
- PyTextRank:
  https://github.com/ceteri/pytextrank
- TensorFlow TextSum (state-of-the-art abstractive summarisation):
  https://github.com/tensorflow/models/tree/master/research/textsum

RESOURCES

- Resoomer, text summarisation demo: https://resoomer.com/en/



ASSOCIATED READING

- Jurafsky, Daniel, and James H. Martin. 2009. Speech and Language Processing: An Introduction to Natural Language Processing, Speech Recognition, and Computational Linguistics. **Chapter 23.**