NLP Systems & Applications: Introduction

LING 573 NLP Systems and Applications Ryan Georgi, March 27, 2018







Roadmap

Motivation

- Course Structure
- Shared Tasks
- Summarization: An Introduction







Motivation: Why Summarization?







Motivation

- Information Retrieval is Very Powerful
 - Search engines index and search enormous doc sets
 - Retrieve billions of documents in tenths of seconds
- But still very limited!
 - Technically: keyword search (mostly)
 - Conceptually: User seeks information
 - Sometimes a website or document
 - Sometimes the answer to a question
 - Sometimes, a summary of a document or document set









- Even web search relies on simple summarization
 - Provide "thumbnail" of relevant information in ranked document.

Caldera - Wikipedia

https://en.wikipedia.org/wiki/Caldera A *caldera* is a large cauldron-like depression that forms following the evacuation of a magma chamber/reservoir. Yellowstone Caldera · Bárðarbunga · Cauldron · Long Valley Caldera

Caldera: Crater Formed by Volcanic Collapse or Explosion

https://geology.com/articles/caldera/ They are large volcanic craters that form by two different methods: 1) an explosive volcanic eruption; or, 2) collapse of surface rock into an empty magma chamber. The accompanying image is a satellite view of one of the most famous *calderas* - Crater Lake in Oregon.

How Volcanoes Work - Calderas - SDSU geology

www.geology.sdsu.edu/how_volcanoes_work/Calderas.html CALDERAS. When an erupting volcano empties a shallow-level magma chamber, the edifice of the volcano may collapse into the voided reservoir, thus forming a steep, bowl-shaped depression called a caldera (Spanish for kettle or cauldron). These features are highly variable in size, ranging from 1-100 km in diameter.



Why Summarization?







- Complex questions go beyond factoids & infoboxes
 - Require explanations, analysis
- For instance: Is acetaminophen or ibuprofen better for reducing fever in kids?
- Following is top-ranked page:











WASHINGTON

Acetaminophen or lbuprofen? top hit: parents.com/health/hygiene/childrens-health-myths/

Ibuprofen beats acetaminophen for treating both pain and fever, according to recent research. One large study in the Archives of Pediatrics & Adolescent Medicine found that ibuprofen lowered kids' fevers better than acetaminophen at two, four, and six hours after taking the medicine. Another study of kids in the ER concluded that ibuprofen provided significantly better pain relief than acetaminophen (or codeine) for broken bones, bruises, and sprains. Ibuprofen and acetaminophen both act on pain receptors in the brain, but ibuprofen also has an anti-inflammatory effect that helps decrease swelling. "That may also make it a better choice for ear infections, which typically involve inflammation," says Richard Rosenfeld, MD, a pediatric ear, nose, and throat specialist at Long Island College Hospital, in Brooklyn. Ibuprofen also lasts longer than acetaminophen, making it more likely that your child will sleep through the night, especially in the early stages of an ear infection. Interestingly, a new study



4 tsp -	20 mL
3 tsp	15 mL
2 tsp	10 mL
1 tsp	5 mL

A	ceta	mi	NO	ph
top	hit: pare	ents.co	om/h	<u>ealth</u>

- Summary:
 - to recent research.

4 tsp 💷	20 mL
3 tsp -	15 mL
2 tsp	10 mL
1 tsp -	5 mL
	-



en or lbuprofen? <u>h/hygiene/childrens-health-myths/</u>

• Ibuprofen beats acetaminophen for treating both pain and fever, according







- Huge scale, explosive growth in online content
 - 2-4K articles in PubMed daily, 41.7M articles/mo on WordPress alone (2014)
 - How can we manage it?
 - Lots of aggregation sites
 - Effective summarization rarer
- Recordings of meetings, classes MOOCs
 - Slow to access linearly, awkward to jump around
 - Structured summary can be useful.
 - Outline of how-tos, to-dos, etc.







• As <u>Torres-Moreno (2014</u>) puts it:

• "too much information kills information"









- This is a topic that's been under research for half a century
 - ...but we're trying to cover the topic in 10 weeks!

- Another thought:
 - With 30+ students, my lesson plan is fairly broad by necessity of varied backgrounds • ...what if an automatic summarizer could weight your individual priors?







Perspectives on Summarization

● <u>DUC, TAC</u> (2001−...):

- Single- & multi-document summarization
- Readable, concise summaries
- Largely news-oriented
 - Later, blogs, etc also query-focused
- Text simplification
 - Compress, simplify text for enhanced readability



Application to CALL, reading levels (e.g. Simple Wikipedia), assistive technology





NLP and Summarization

- Rich testbed for NLP techniques:
 - Information Retrieval
 - Named Entity Recognition
 - Word, sentence segmentation
 - Information extraction
 - Parsing
 - Semantics
 - Discourse relations
 - Co-reference
 - Generation
 - Paraphrasing

Deep/shallow techniques; machine learning

WASHINGTON



PROFESSIONAL MASTER'S IN COMPUTATIONAL LINGUISTICS



Course Structure



PROFESSIONAL MASTER'S IN COMPUTATIONAL LINGUISTICS



573 Structure

• Implementation:

- Create a summarization system
 - Extend existing software components
 - Develop, evaluate on standard data set

• **Report + Presentation**:

- Write a technical report
- Present plan, system, results in class
- Give/receive feedback







Implementation: Deliverables

- Complex system
 - Break into (relatively) manageable components
 - Incremental progress, deadlines
- Key components
 - **DI**: Setup
 - **D2**: Baseline System, Content Selection
 - **D3**: Content Selection, Information Ordering
 - **D4**: Content selection, Information Ordering, Surface Realization, Final Results
- Deadlines:
 - Little slack in schedule; please keep to deadlines
 - Timing: ~12 hours/wk; sometimes higher





• Technical report:

- Follow organization for scientific paper
- Both formatting and content

• Presentations

- 10-15 minute oral presentation for deliverables
- Explain goals, methodology, success, issues
- Critique each others' work
- Attend **ALL** presentations



Report + Presentation





Working In Teams

- Why teams?
 - Representative of professional environment
 - Too much work for a single person
- Team organization:
 - Form groups of 3 (possibly 4) people
 - Arrange coordination
 - Distribute work equitably
 - All team members receive the same base grade
 - End-of-course team evaluation
 - Self-and-teammate evaluation
 - Grades may be adjusted in case of severe imbalance







First Task

- Form teams:

 - Also a discussion thread to find other groupmates [link]
- Decide upon tooling
 - (Private) repository location
 - Communication Methods / Task Tracking



• Post to the team formation discussion thread on Canvas with your groupmates: [link]





- Readings:
 - Current research papers in summarization [course readings page]
 - Jurafsky & Martin; Manning & Schütze
 - Background, reference, refresher
- Software:
 - Build on existing system components, toolkits
 - NLP, ML Toolkits
 - Corpora



Resources





Resources: Patas

- All code must run on Patas
- Also available on patas:
 - Software systems: /NLP_TOOLS
 - Corpora: /corpora
 - Repositories
 - SVN (Contact Brandon)
 - Git via:
 - GitLab (request-git-account)
 - SSH (Contact Brandon)

PROFESSIONAL MASTER'S IN COMPUTATIONAL LINGUISTICS

Shared Tasks

Shared Tasks: Goals

- Focus research community on key challenges
- Support the creation of large-scale community resources
 - Corpora: News, Recordings, Video
 - Annotation: Expert questions, labeled answers
- Develop methodologies to evaluate state-of-the-art
 - Retrieval (MRR), Machine Translation (BLEU), etc
- Facilitate technology/knowledge transfer between industry & academia

Shared Tasks: Goals

Pragmatic

- Head-to-head comparison of systems & techniques
 - Same data, same task, same conditions, same timing
- Centralizes funding, effort
- Requires disclosure of techniques in exchange for data

Base

- Bragging rights
- Government research funding decisions

Shared Tasks: Perspective

- Late '80s-90s
 - ATIS: spoken dialog systems
 - MUC: Message Understanding (information extraction)
- TREC (Text REtreival Conference)
 - Arguably largest (often >100 participating teams)
 - Longest running (1992–current)
 - Information retrieval (and related technologies)
 - Actually hadn't had 'ad-hoc' since 2000-2016
 - Organized by NIST

TREC Tracks

- "Track" Basic task organization
- Previous tracks:
 - Ad-hoc Basic retrieval from fixed document set
 - Cross-language Query in one language, document set in another
- English, French, Spanish, Italian, German, Chinese, Arabic
 - Genomics
 - Spoken Document Retrieval
 - Video Search
 - Question Answering

Other Shared Tasks

- International:
 - <u>CLEF</u> (Europe); <u>FIRE</u> (India)
- Other NIST
 - Machine Translation
 - Topic Detection & Tracking
- Various:
 - <u>CoNLL</u> (NER, Parsing, +...)
 - <u>SemEval</u> (WSD)
 - PASCAL (morphology, grammar induction, textual entailment)

• ...& more

Summarization History

- "The Automatic Creation of Literature Abstracts"
 - Luhn, 1956 Early IBM system based on word, sentence statistics
- 1993 Dagstuhl Seminar:
 - Meeting launched renewed interest in summarization
- 1997 ACL Summarization Workshop

Summarization Campaigns

<u>SUMMAC</u> (1998)

- Initial cross-system evaluation campaign
- <u>DUC</u> (Document Understanding Conference) (2001–2007)
 - Increasing complexity, including multi-document, topic-oriented, multi-lingual tracks Developed systems and evaluation in tandem
- <u>NTCIR</u> (3 Years)
 - Single, multi-document; Japanese

Most Recent Summarization Campaigns

- <u>TAC</u> (Text Analytics Conference): 2008–Current
 - Summarization systems:
 - Opinion
 - Update
 - Guided
 - Multilingual
- - Three editions to date
 - Scientific document summarization
 - Facets and citations

WASHINGTON

<u>CL-SCISUMM</u>: Computational Linguistics SClentific document SUMmarization shared task

Summarization Tasks

• Provide

- List of topics (e.g. "guided" summarization)
- Document collections (licensed via LDC, NIST)
- Lists of relevant documents
- Validation tools
- Evaluation tools: Model summaries, systems
- Derived resources:
 - Baseline systems, preprocessing tools, components
- Reams of related publications

Topics

<topic id = "D0906B" category = "1"> <docsetA id = "D0906B-A"> <doc id = "AFP ENG 20050110.0079" /> <doc id = "LTW ENG 20050110.0006" /> <doc id = "LTW ENG 20050112.0156" /> <doc id = "NYT ENG 20050110.0340" /> <doc id = "NYT ENG 20050111.0349" /> <doc id = "LTW ENG 20050109.0001" /> <doc id = "LTW ENG 20050110.0118" /> <doc id = "NYT ENG 20050110.0009" /> <doc id = "NYT ENG 20050111.0015" /> <doc id = "NYT ENG 20050112.0012" /> </docset> <docsetB id = "D0906B-B"> <doc id = "AFP ENG 20050221.0700" />

WASHINGTON

<title> Rains and mudslides in Southern California </title>

Documents

<DOC><DOCNO> APW20000817.0002 </DOCNO>
<DOCTYPE> NEWS STORY </DOCTYPE><DATE_TIME> 2000-08-17 00:05 </
DATE_TIME>

<BODY> <HEADLINE> 19 charged with drug trafficking </HEADLINE>
<TEXT><P>

UTICA, N.Y. (AP) - Nineteen people involved in a drug trafficking ring in the Utica area were arrested early Wednesday, police said. </P><P>

Those arrested are linked to 22 others picked up in May and comprise ''a major cocaine, crack cocaine and marijuana distribution organization,'' according to the U.S. Department of Justice. </P>

Model Summaries

<SUM>

<aid="1.2">In January 2005</aid="1.2">, <aid="1.7">rescue workers

<aid="1.3">in southern California</aid="1.3"> used snowplows, snowcats and snowmobiles to free <aid="1.5">people</aid="1.5">

from a highway where

</aid="1.7">

<aid="1.1">snow, sleet, rain and fog caused a 200-vehicle logjam</aid="1.1">. <aid="1.1">A fourth day of storms took a heavy toll as saturated hillsides gave way aid="1.1">,

<aid="1.6">mudslides inundating houses and closing highways</aid="1.6">. <aid="1.5">People fled neighborhoods up and down the coast.</aid="1.5"> Eight of nine horse races at Santa Anita were canceled for the first time in 10 years. <aid="1.6">More than 6,000 houses were without power</aid="1.6"> <aid="1.3">in Los Angeles</aid="1.3">.

A scientist said Los Angeles had not seen such intensity of winter downpours since 1889-90.

</SUM>

Summarization: An Introduction

The Summarization Task

- Has long been a market for humansummarized texts
 - Reader's Digest Condensed Books
 - Now: Blinkist, BookShots

- ...every academic paper does this as an abstract!
- What does the human process look like?
 - "Understanding" and "Selection"
 - ...that's a whole lot of handwaving to automate!

Structuring the Summarization Task

- Summarization task definition via <u>Mani and Maybury, 1999</u>
 - for a particular task and user
- Main components:
 - Content selection
 - Information ordering
 - Sentence realization

• Process of distilling the most important information from a text to produce an abridged version

Dimensions of Summarization

- Rich problem domain
- Tasks and systems vary on:
 - Use purpose
 - Audience
 - Derivation
 - Coverage
 - Reduction
 - Input/Output form factors

PROFESSIONAL MASTER'S IN COMPUTATIONAL LINGUISTICS

Dimensions of Summarization: Purpose

- What is the goal of the summary? How will it be used?
 - Often this is surprisingly vague
 - Generic "reflective" summaries: highlight prominent content
 - Relevance filtering: "Indicative" quickly tell if document covers desired content
 - Browsing, skimming
 - Compression for assistive tech
 - Briefings medical summaries, to-do lists; definition Q/A

Dimensions of Summarization: Audience

- Experts vs. Novice/Generalists
- News summaries
 - "ordinary" reader or analyst?
 - Many funded evaluation programs target analysts
- Medical
 - Patient-directed?
 - Practicioner-directed?
 - Generalist or specialist?

PROFESSIONAL MASTER'S IN COMPUTATIONAL LINGUISTICS

Dimensions of Summarization: Derivation

- How to generate the summaries
 - **Extractive** Built from units extracted from original text
 - Abstractive Concepts from source, generate new text as output

Extract vs. Abstract

• **Extract** from the Gettysburg Address

• Abstract of the Gettysburg Address (Mani, 2001)

Four score and seven years ago our fathers brought forth upon this content a new nation, conceived in liberty, and dedicated to the proposition that all men are created equal. Now we are engaged in a great civil war, testing whether that nation can long endure. We are met on a great battlefield of that war. We have come to dedicate a portion of that field. But the brave men, living and dead, who struggled here, have consecrated it far above our poor power to add or detract. From these honored dead we take increased devotion to that cause for which they gave the last full measure of devotion — that government of the people, by the people, and for the people shall not perish from the earth.

The speech by Abraham Lincoln commemorates soldiers who laid down their lives in the Battle of Gettysburg. It reminds the troops that it is the future of freedom in America that they are fighting for.

Dimensions of Summarization: Coverage, Units, Reduction

Coverage

• Comprehensive vs. Query/Topic Oriented?

Units

- Single vs. Multi-Document
- **Reduction** (aka compression)
 - Percentage (50%) or absolute length (max 100 words).

PROFESSIONAL MASTER'S IN COMPUTATIONAL LINGUISTICS

Dimensions of Summarization: Input/Output Form Factors

- Language
 - Single language? Multiple languages, with translation and/or semantic analysis?
- Register Formality, or style
- Genre News, sports, medical, technical...
- Structure Forms, tables, lists, web pages
- Medium Text, speech, images, video, tables

Dimensions of Summarization: Evaluation

- Summary evaluation is very hard!
 - Using human-generated abstracts, surprisingly little overlap (substantial assessor disagreement)
 - Usually developed in parallel with the particular system/task
- Key concepts:

- Text quality readability (sentence & discourse structure)
- Concept capture are key concepts covered?
- Gold standards human summaries for comparison, incorporation of specific goals
- Purpose why is the summary created?

General Architecture

PROFESSIONAL MASTER'S IN COMPUTATIONAL LINGUISTICS

Next Time

- Continued Overview:
 - A broader overview of the summarization task.
- Evaluation:
 - What makes a good/bad summarization?

Office Hours!

- If you have a preference for the TA office hours, please fill out the poll at:
 - https://doodle.com/poll/3efue9gtndmctm73
- Mine will be 12:30—2:30 on Wednesdays.

