Systems & Applications: Introduction

Ling 573
NLP Systems and Applications
March 28, 2017

Roadmap

- Motivation
- 573 Structure
- Summarization
- Shared Tasks

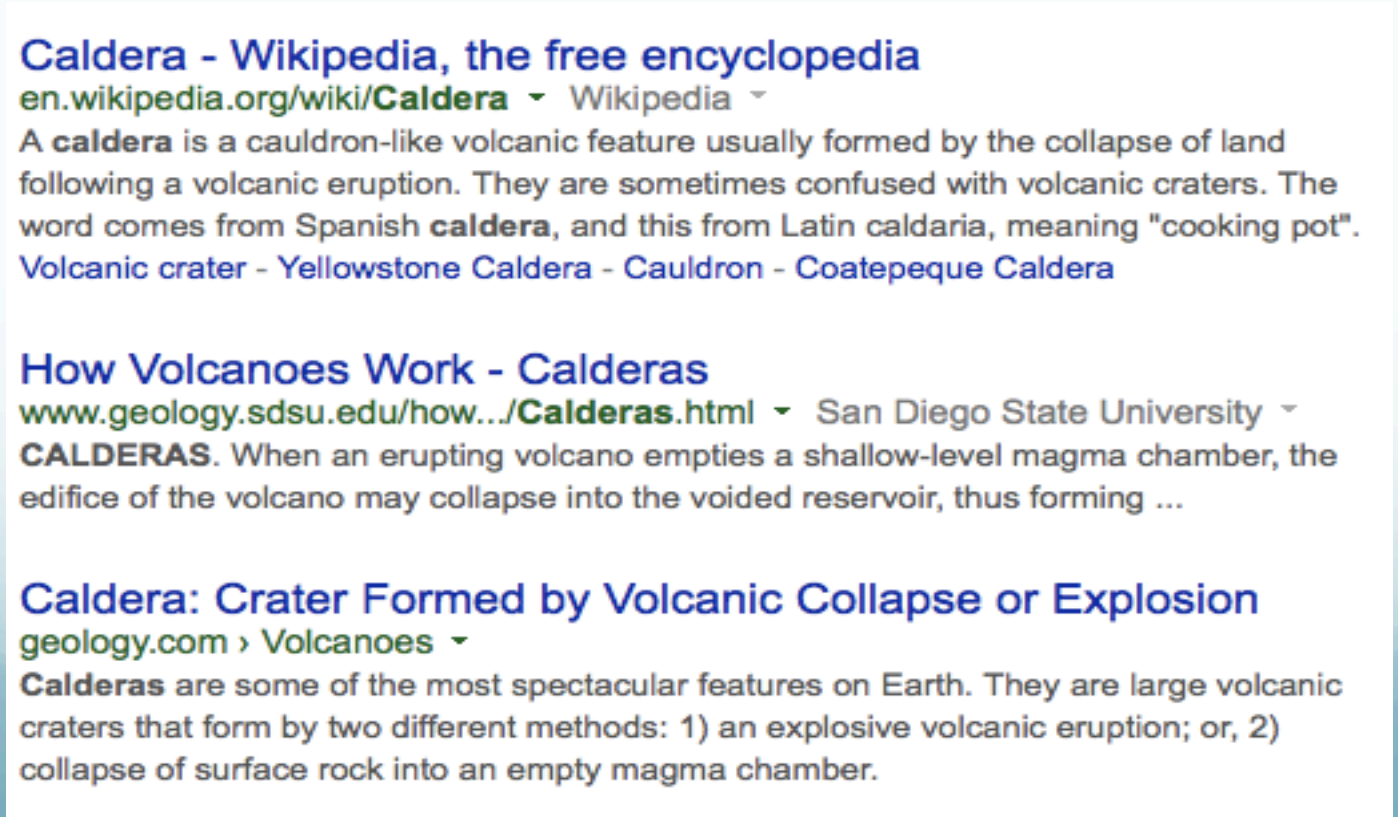
Motivation

- Information retrieval is very powerful
  - Search engines index and search enormous doc sets
  - Retrieve billions of documents in tenths of seconds

- But still limited!
  - Technically – keyword search (mostly)
  - Conceptually
    - User seeks information
      - Sometimes a web site or document
      - Sometimes the answer to a question
      - But, often a summary of document or document set

Why Summarization?

- Even web search relies on simple summarization
  - Snippets!
    - Provide thumbnail summary of ranked document



Why Summarization?

- Complex questions go beyond factoids, infoboxes
  - Require explanations, analysis
    - E.g. Is acetaminophen or ibuprofen better for reducing fever in kids?
    - Highest search hit is parenting page
      - Provides a multi-document summary

http://www.parents.com/health/hygiene/childrens-health-myths/#page=1

Ted Morrison

Acetaminophen or Ibuprofen?

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 also found that children who took acetaminophen before age 1 were almost 50 percent more likely to

Why Summarization?

- Complex questions go beyond factoids, infoboxes
  - Require explanations, analysis
    - E.g. Is acetaminophen or ibuprofen better for reducing fever in kids?
    - Summary: Ibuprofen beats acetaminophen for treating both pain and fever, according to recent research.

Why Summarization?

- 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,

Perspectives on Summarization

- DUC, TAC (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
      - Also aims to support greater automation

Natural Language Processing and Summarization

- Rich testbed for NLP techniques:
  - Information retrieval
  - Named Entity Recognition
  - Word, sentence segmentation
  - Information extraction
  - Parsing
  - Semantics, etc..
  - Discourse relations
  - Co-reference
  - Generation
  - Paraphrasing
- Deep/shallow techniques; machine learning

573 Structure

- Implementation:
  - Create a summarization system
    - Extend existing software components
    - Develop, evaluate on standard data set
- 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:
  - D1: 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 time
  - Timing: ~12 hours week; sometimes higher

Presentation

- Technical report:
  - Follow organization for scientific paper
    - Formatting and Content

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

Working in Teams

- Why teams?
  - Too much work for a single person
  - Representative of professional environment
- Team organization:
  - Form groups of 3 (possibly 2) 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:
  - Email Glenn gslayden@uw.edu with the team list

Resources

- Readings:
  - Current research papers in summarization
  - Jurafsky & Martin/Manning & Schutze text
    - Background, reference, refresher

- Software:
  - Build on existing system components, toolkits
    - NLP, machine learning, etc
    - Corpora, etc

Resources: Patas

- System should run on patas
  - Existing infrastructure
    - Software systems
    - Corpora
    - Repositories

Shared Task Evaluations

- Goals:
  - Lofty:
    - Focus research community on key challenges
      - 'Grand 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, Machine Translation, etc
    - Facilitate technology/knowledge transfer b/t industry/acad.

Shared Task Evaluation

- 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 Retrieval 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, though
  - 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, docs in other
    - English, French, Spanish, Italian, German, Chinese, Arabic
  - Genomics
  - Spoken Document Retrieval
  - Video search
  - Question Answering

Other Shared Tasks

- International:
  - CLEF (Europe); FIRE (India)
- Other NIST:
  - Machine Translation
  - Topic Detection & Tracking
- Various:
  - CoNLL (NE, parsing,..); SENSEVAL: WSD; PASCAL (morphology); BioNLP (biological entities, relations)
  - Mediaeval (multi-media information access)

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

- SUMMAC: (1998)
  - Initial cross-system evaluation campaign
- DUC (Document Understanding Conference)
  - 2001-2007
    - Increasing complexity, including multi-document, topic-oriented, multi-lingual
    - Developed systems and evaluation in tandem
- NTCIR (3 years)
  - Single, multi-document; Japanese

Most Recent Summarization Campaigns

- TAC (Text Analytics Conference): 2008---current
  - Variety of tasks
    - Summarization systems:
      - Opinion
      - Update
      - Guided
      - Multi-lingual
    - Automatic evaluation methodology
- CL-SCISUMM: three editions to date
  - Scientific document summarization
    - Facets and citations

Summarization Tasks

- Provide:
  - Lists 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, pre-processing tools, components
  - Reams of related publications

Topics

- <topic id = "D0906B" category = "1">
  - <title> Rains and mudslides in Southern California </title>
    - <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" />
      - .....

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>

Structuring the Summarization Task

- Summarization Task: (Mani and Mayberry 1999)
  - Process of distilling the most important information from a text to produce an abridged version for a particular task and user

- Main components:
  - Content selection
  - Information ordering
  - Sentence realization

Dimensions of Summarization

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

Dimensions of Summarization

- Purpose:
  - What is the goal of the summary? How will it be used?
    - Often 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:
  - Who is the summary for?
    - Also related to the content
    - Often contrasts experts vs novice/generalists
  - News summaries:
    - ‘Ordinary’ vs analysts
      - Many funded evaluation programs target analysts
  - Medical:
    - Patient directed vs doctor/scientist-directed

Dimensions of Summarization

- "Derivation":
  - Continuum
    - Extractive: Built from units extracted from original text
    - Abstractive: Concepts from source, generated in final form
  - Predominantly extractive
- Coverage:
  - Comprehensive (generic) vs query-/topic-oriented
    - Most evaluations focused
- Units: single vs multi-document
- Reduction (aka compression):
  - Typically percentage or absolute length

Extract vs Abstract

Extract from the Gettysburg Address:
Four score and seven years ago our fathers brought forth upon this continent 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 for the people shall not perish from the earth.
Abstract of the Gettysburg Address:
This 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.

Figure 23.13 An extract versus an abstract from the Gettysburg Address (abstract from Mani (2001)).

Dimensions of Summarization

- Input/Output form factors:
  - Language: Evaluations include:
    - English, Arabic, Chinese, Japanese, multilingual
  - Register: Formality, style
  - Genre: e.g. News, sports, medical, technical,....
  - Structure: forms, tables, lists, web pages
  - Medium: text, speech, video, tables
  - Subject