Summarization
Machine Translation

Summarization

- *Text summarization is the process of distilling the most important information from a text to produce an abridged version for a particular task and user*
  - Definition adapted from Mani and Maybury 1999
- Types of summaries in current research:
  - Outlines of any document
  - Abstracts of a scientific article
  - Headlines of a news article
  - Snippets summarizing a Web page or a search engine results page
  - Action items or other summaries of a business meeting
  - Summaries of email threads
  - Compressed sentences for simplified or clarified text
  - Single-document vs. multi-document summarization

Typical approaches to general problem

- Currently, a true re-phrasing overall summary is not yet achievable. Most systems primarily select sentences from a document and do some rephrasing
  - **Content Selection**
    - Identify the sentences of clauses to extract
  - **Information Ordering**
    - How to order the selected units
  - **Sentence Realization**
    - Perform cleanup on the extracted units so that they are fluent in their new context.

Content Selection

- Simple approach is to select sentences that have more informative words according to saliency defined from a topic signature of the document
- Centroid-based summarization uses log-likelihood ratios for words, computing the probability of observing the word in the input more often than in the background corpus
- Other centrality methods try to rank the sentences according to a centrality score
- Methods based on rhetorical parsing use coherence relations to identify satellite and nuclues sentences
- Machine learning methods use features based on
  - Position, cue phrases, word informativeness, sentence length, cohesion (computing lexical chains of the document)

Summarization Evaluation

- Extrinsic (task-based) evaluation: humans are asked to rate the summaries according to how well they are enabled to perform a specific task
- Intrinsic (task-independent) evaluation
  - Human judgments to rate the summaries
  - ROUGE
    - Humans generate summaries for a document collection
    - System-generated summaries are rated according to how close they come to the human-generated summary
    - Measures have included unigram overlap, bigram overlap, and longest common subsequence
  - Pyramid method
    - Humans identify "units of meaning" and then an overlap measure is computed

Machine Translation

- Translating text from one language to another is a task challenging even for humans to try to fully capture the style and nuanced meaning of the original
- While research focuses on trying to produce the fully-automatic, high-quality translation, there are many tasks for which a rough translation is sufficient
- The differences between languages include systematic differences that can be modeled in some way and idiosyncratic and lexical differences that must be dealt with one by one.

Why MT is hard

- Given the Japanese phrase
  fukaku hansei shite orimasu
- If this is translated to English as
  we apologize
  it is not faithful to the original meaning
- But if we translate it as
  we are deeply reflecting (on our past behavior, and what we did wrong, and how to avoid the problem next time)

  the translation is not fluent.

  Example from Jurafsky and Martin text.

Classical MT

- In this line of MT research, approaches can be classified according to the level of unit of translation
  - See the Vauquois triangle
  - Direct translation uses a word translation approach
  - Syntactic and semantic transfer approaches use syntactic phrase and semantic units, respectively, as the unit of translation

Statistical Approaches

- Build probabilistic models of faithfulness and fluency and combine the models to produce the most probable translation.
- Modeled as a noisy channel "pretend that the foreign input F is a corrupted version of the target language output E and the task is to discover the hidden sentence E that generated the observed sentence F."
- Requires three components
  - Language model to compute P(E)
  - Translation model to compute P(F|E)
  - Decoder, which is given F and produces the most probable E
    - Usually phrase-based

Alignment and Parallel Corpora

- All translation models are based on probabilities of word alignment
- Word alignment models are automatically trained from parallel corpora
- Hansard corpora work best for this
  - Translations of official government documents
    - Canadian parliament documents for French, English and a variety of native American languages
    - United Nations proceedings documents
  - Literary parallel corpora are not as suitable because of the stronger presence of literary devices, such as metaphor

MT Evaluation

- Human raters can evaluate along the two dimensions of fluency and fidelity (and there are several individual metrics for each of these dimensions)
- BLEU automatic evaluation system
  - Evaluation corpus contains human generated translations
  - Metrics evaluate how closely the system-generated translations correspond to the human ones