# *Optimization Based Approaches for Summarization*

Pawan Goyal

CSE, IIT Kharagpur

Week 11, Lecture 2

# Global Inference

- Let us define document $D$ with $t_n$ textual units

$$D = t_1, t_2, \ldots, t_{n-1}, t_n$$

- Let $Rel(i)$ be the relevance of $t_i$ to be in the summary
- Let $Red(i,j)$ be the redundancy between $t_i$ and $t_j$
- Let $l(i)$ be the length of $t_i$

# Inference Problem

- The inference problem is to select a subset $S$ of textual units from $D$ such that summary score of $S$, i.e., $s(S)$, is maximized.
- $S = \arg\max_{S \subseteq D} \left[ \sum_{t_i \in S} Rel(i) - \sum_{t_i, t_j \in S, i<j} Red(i,j) \right]$ such that $\sum_{t_i \in S} l(i) \leq K$, where $k$ denotes the maximum length of the summary

# A Greedy Solution

1. Sort $D$ so that $Rel(i) > Rel(i+1) \forall i$
2. $S = \{t_1\}$
3. while $\sum_{t_i \in S} l(i) < K$
4. $t_j = \arg\max_{t_j \in D-S} s(S \cup \{t_j\})$
5. $S = S \cup \{t_j\}$
6. return $S$

# *Integer Linear Programming (ILP)*

- Greedy algorithm is an approximate solution
- Use exact solution algorithms with ILP
- ILP is a constrained optimization problem
- Many solvers on the web
- Define the constraints based on relevance and redundancy for summarization

# Sentence Level ILP Formulation

## Optimization Function

maximize $\sum_i \alpha_i Rel(i) - \sum_{i<j} \alpha_{ij} Red(i,j)$

# Sentence Level ILP Formulation

## Optimization Function

maximize $\sum_i \alpha_i Rel(i) - \sum_{i<j} \alpha_{ij} Red(i,j)$

## Constraints

such that $\forall i,j$:

- $\alpha_i, \alpha_{ij} \in \{0,1\}$
- $\sum_i \alpha_i l(i) \leq K$
- $\alpha_{ij} - \alpha_i \leq 0$
- $\alpha_{ij} - \alpha_j \leq 0$
- $\alpha_i + \alpha_j - \alpha_{ij} \leq 1$

# Sentence Level ILP Formulation

## Optimization Function

maximize $\sum_i \alpha_i Rel(i) - \sum_{i<j} \alpha_{ij} Red(i,j)$

## Constraints

such that $\forall i,j$:

- $\alpha_i, \alpha_{ij} \in \{0,1\}$
- $\sum_i \alpha_i l(i) \leq K$
- $\alpha_{ij} - \alpha_i \leq 0$
- $\alpha_{ij} - \alpha_j \leq 0$
- $\alpha_i + \alpha_j - \alpha_{ij} \leq 1$

## Is generic enough

Depending on your task, you can define your own optimization function and constrains.

# The next steps: Sentence Ordering

# The next steps: Sentence Ordering

## Chronological ordering: the simplest method

List the sentences in the order, they appear in the document

# The next steps: Sentence Ordering

## Chronological ordering: the simplest method

List the sentences in the order, they appear in the document

## Coherence

- Choose orderings that make neighboring sentences similar (by cosine)
- Choose orderings in which neighboring sentences discuss the same entity

# The next steps: Sentence Ordering

## Chronological ordering: the simplest method

List the sentences in the order, they appear in the document

## Coherence

- Choose orderings that make neighboring sentences similar (by cosine)
- Choose orderings in which neighboring sentences discuss the same entity

## Topical ordering

Learn the ordering of topics in the source documents

# The next steps: Simplifying Sentences

Parse sentences, use rules to decide which modifiers to prune

# The next steps: Simplifying Sentences

Parse sentences, use rules to decide which modifiers to prune

- **Initial adverbials:** *For example, on the other hand, as a matter of fact, at this point, ...*

# The next steps: Simplifying Sentences

Parse sentences, use rules to decide which modifiers to prune

- **Initial adverbials:** *For example, on the other hand, as a matter of fact, at this point, ...*
- **PPs without named entities:** The commercial fishing restrictions in Washington will not be lifted unless the salmon population increases [*PP to a sustainable number*]

# The next steps: Simplifying Sentences

Parse sentences, use rules to decide which modifiers to prune

- **Initial adverbials:** *For example, on the other hand, as a matter of fact, at this point, ...*
- **PPs without named entities:** The commercial fishing restrictions in Washington will not be lifted unless the salmon population increases [*PP to a sustainable number*]
- **Attribution clauses:** Rebels agreed to talks with government officials, *international observers said Tuesday*

# *The next steps: Simplifying Sentences*

Parse sentences, use rules to decide which modifiers to prune

- **Initial adverbials:** *For example, on the other hand, as a matter of fact, at this point, ...*
- **PPs without named entities:** The commercial fishing restrictions in Washington will not be lifted unless the salmon population increases [*PP to a sustainable number*]
- **Attribution clauses:** Rebels agreed to talks with government officials, *international observers said Tuesday*
- **Appositives:** Rajan*, 28, an artist who was living at the time in Philadelphia,* found the inspiration in the back of city magazines