

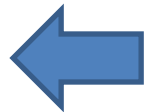
Multi-Document Summarization By Sentence Extraction

Authors: Jade Goldstein, Vibhu Mittal,
Jaime Carbonell, Mark Kantrowitz

Presented by: Xiaoqian Liu

Previous work

- Text-span deletion (@IBM, 1958)
- Sentence extraction
 - Statistical techniques (e.g. frequency analysis)
 - Information extraction + language generation
 - Pattern recognition by machine learning
- Extend single document summarization techniques
 - Challenges



Motivation

- Use domain-independent techniques only
- Reduce redundancy without eliminating potentially relevant information
- Easy Parameterization

Requirements

- **Clustering**
- **Coverage**
- **Anti-redundancy**
- Effective user interfaces
- Context
- Summary Update
- Coherence
- Identification of source inconsistencies

Multi-document summarizer

- Segment the documents into passages, and index them using inverted indices
- Identify the passages relevant to the query based on cosine similarity
- Apply the MMR-MD metric. Given with the length of the summary, select a number of passages
- Create a summary over the selected passages

Maximal Marginal Relevance – Multiple Documents (MMR-MD)

- Maximize *relevant novelty*
 - Linear combination of relevance and novelty
 - i.e. “Marginal Relevance”

$MMR - MD$

$$\stackrel{\text{def}}{=} \text{Arg} \max_{P \in R \setminus S} [\lambda Sim_1 - (1 - \lambda) \max_{P \in S} Sim_2]$$

Sim_1 : Metric for relevance ranking (between query and document)

Sim_2 : Anti – redundancy metric (between passages)

λ : scoring factor

MMR-MD Algorithm

$$MMR-MD \stackrel{\text{def}}{=} \text{Arg} \max_{P_{ij} \in R \setminus S} \left[\lambda (\text{Sim}_1(P_{ij}, Q, C_{ij}, D_i, D)) - (1 - \lambda) \max_{P_{nm} \in S} \text{Sim}_2(P_{ij}, P_{nm}, C, S, D_i) \right]$$

$$\text{Sim}_1(P_{ij}, Q, C_{ij}, D_i, D) = w_1 * (P_{ij} \cdot Q) + w_2 * \text{coverage}(P_{ij}, C_{ij}) + w_3 * \text{content}(P_{ij}) + w_4 * \text{time_sequence}(D_i, D)$$

$$\text{Sim}_2(P_{ij}, P_{nm}, C, S, D_i) = w_a * (P_{ij} \cdot P_{nm}) + w_b * \text{clusters_selected}(C_{ij}, S) + w_c * \text{documents_selected}(D_i, S)$$

$$\text{coverage}(P_{ij}, C) = \sum_{k \in C_{ij}} w_k * |k|$$

$$\text{content}(P_{ij}) = \sum_{W \in P_{ij}} w_{\text{type}}(W)$$

$$\text{time_sequence}(D_i, D) = \frac{\text{timestamp}(D_{\text{maxtime}}) - \text{timestamp}(D_i)}{\text{timestamp}(D_{\text{maxtime}}) - \text{timestamp}(D_{\text{mintime}})}$$

$$\text{clusters_selected}(C_{ij}, S) = |C_{ij} \cap \bigcup_{v, w: P_{vw} \in S} C_{vw}|$$

$$\text{documents_selected}(D_i, S) = \frac{1}{|D_i|} * \sum_w [P_{iw} \in S]$$

Experiment

- TIPSER evaluation corpus -> query
- News documents from AP & Wall Street Journal:
 - 6115 sentences, 200 documents
- Sentence as summary unit
- 10 sentence-long summary
 - Sentence compression ratio: 0.2%
 - Character compression ratio: 0.3%
- $\lambda = 0.3$ and 1

$\lambda = 1$ (Relevance Ranking)

1. **WSJ910204-0176: 1** CAPE TOWN, South Africa – President F.W. de Klerk’s proposal to repeal the major pillars of apartheid drew a generally positive response from black leaders, but African National Congress leader Nelson Mandela called on the international community to continue economic sanctions against South Africa until the government takes further steps.
2. **AP880803-0082: 25** Three Canadian anti-apartheid groups issued a statement urging the government to sever diplomatic and economic links with South Africa and aid the African National Congress, the banned group fighting the white-dominated government in South Africa.
3. **AP880803-0080: 25** Three Canadian anti-apartheid groups issued a statement urging the government to sever diplomatic and economic links with South Africa and aid the African National Congress, the banned group fighting the white-dominated government in South Africa.
4. **AP880802-0165: 23** South Africa says the ANC, the main black group fighting to overthrow South Africa’s white government, has seven major military bases in Angola, and the Pretoria government wants those bases closed down.
5. **AP880212-0060: 14** ANGOP quoted the Angolan statement as saying the main causes of conflict in the region are South Africa’s “illegal occupation” of Namibia, South African attacks against its black-ruled neighbors and its alleged creation of armed groups to carry out “terrorist activities” in those countries, and the denial of political rights to the black majority in South Africa.
6. **AP880823-0069: 17** The ANC is the main guerrilla group fighting to overthrow the South African government and end apartheid, the system of racial segregation in which South Africa’s black majority has no vote in national affairs.
7. **AP880803-0158: 26** South Africa says the ANC, the main black group fighting to overthrow South Africa’s white-led government, has seven major military bases in Angola, and it wants those bases closed down.
8. **AP880613-0126: 15** The ANC is fighting to topple the South African government and its policy of apartheid, under which the nation’s 26 million blacks have no voice in national affairs and the 5 million whites control the economy and dominate government.
9. **AP880212-0060: 13** The African National Congress is the main rebel movement fighting South Africa’s white-led government and SWAPO is a black guerrilla group fighting for independence for Namibia, which is administered by South Africa.
10. **WSJ870129-0051: 1** Secretary of State George Shultz, in a meeting with Oliver Tambo, head of the African National Congress, voiced concerns about Soviet influence on the black South African group and the ANC’s use of violence in the struggle against apartheid.

Almost identical

Figure 2: Sample multi-document summary with $\lambda = 1$, news-story-principle ordering (rank order)

$\lambda = 0.3$ (query relevance, anti-redundancy)

- 
1. **WSJ870129-0051 1** Secretary of State George Shultz, in a meeting with Oliver Tambo, head of the African National Congress, voiced concerns about Soviet influence on the black South African group and the ANC's use of violence in the struggle against apartheid.
 2. **WSJ880422-0133 44** (See related story: "ANC: Apartheid's Foes – The Long Struggle: The ANC Is Banned, But It Is in the Hearts of a Nation's Blacks — In South Africa, the Group Survives Assassinations, Government Crackdowns — The Black, Green and Gold" – WSJ April 22, 1988)
 -  3. **AP880803-0158 26** South Africa says the ANC, the main black group fighting to overthrow South Africa's white-led government, has seven major military bases in Angola, and it wants those bases closed down.
 4. **AP880919-0052 5** But activist clergymen from South Africa said the pontiff should have spoken out more forcefully against their white-minority government's policies of apartheid, under which 26 million blacks have no say in national affairs.
 5. **AP890821-0092 10** Besides ending the emergency and lifting bans on anti-apartheid groups and individual activists, the Harare summit's conditions included the removal of all troops from South Africa's black townships, releasing all political prisoners and ending political trials and executions, and a government commitment to free political discussion.
 6. **WSJ900503-0041 11** Pretoria and the ANC remain far apart on their visions for a post-apartheid South Africa: The ANC wants a simple one-man, one-vote majority rule system, while the government claims that will lead to black domination and insists on constitutional protection of the rights of minorities, including the whites.
 7. **WSJ900807-0037 1** JOHANNESBURG, South Africa – The African National Congress suspended its 30-year armed struggle against the white minority government, clearing the way for the start of negotiations over a new constitution based on black-white power sharing.
 8. **WSJ900924-0119 20** The African National Congress, South Africa's main black liberation group, forged its sanctions strategy as a means of pressuring the government to abandon white-minority rule.
 9. **WSJ910702-0053 36** At a meeting in South Africa this week, the African National Congress, the major black group, is expected to take a tough line against the white-run government.
 -  10. **WSJ910204-0176 1** CAPE TOWN, South Africa – President F.W. de Klerk's proposal to repeal the major pillars of apartheid drew a generally positive response from black leaders, but African National Congress leader Nelson Mandela called on the international community to continue economic sanctions against South Africa until the government takes further steps.

Figure 3: Sample multi-document summary with $\lambda = 0.3$, time-line ordering

Conclusion

- Statistical method of multi-document summarization
- Builds on single-document summarization
- Domain-independent
- Maximize relevant novelty
- Constraint: lack co-reference resolution (need advanced NLP techniques)