**Multi-Document Summarization via the Minimum Dominating Set**

Chao Shenand Tao Li

Summary submitted by *AnunayaSrivastava*

**AIM**

To produce a condensation(or summary) for the given set of documents

**INPUT**

A sentence graph generated from a set of documents where vertices represent sentences and edges indicate thatthe corresponding vertices are similar.

**NOVELTY**

* A new principled and versatile framework for multi-document summarization using the minimum dominatingset is proposed.
* It is shown that four well-knownsummarization tasks including generic,query-focused, update, and comparative summarization can be modeled asdifferent variations derived from theproposed framework.
* Approximationalgorithms for performing summarization are also proposed.

**TYPES OF SUMMARIZATION**

Different types of summarization differ on the basis of their input.

1. Generic Summarization: Input is a set of documents
2. Query-Focused Summarization:

|  |  |  |
| --- | --- | --- |
| Sr. No. | Summarization Type | Input |
| 1 | Generic Summarization | Set of documents |
| 2 | Query-Focused Summarization | Set of documents + User Query ‘q’ |
| 3 | Update Summarization | User query ‘q’ + 2 sets of documents C1, C2 |
| 4 | Comparative Summarization | N group of documents – C1, C2,…, CN |

**HIGH LEVEL APPROACH**

MINIMUM DOMINATING SET

Given a graphG=<V,E>, a dominatingset of Gis a subset Sof vertices with thefollowing property: each vertex of Gis eitherin the dominating setS, or is adjacent to somevertices inS.

**CHALLENGES& APPROXIMATION ALGO**

Finding MDS is an NP-hard problem.

A greedy approximation algorithm for Set Cover(SC) problem is described in Jhonson, 1973. SC problem is also NP-hard and there exists a pair of polynomial time reduction between MDS and SC.

The author suggested a similar greedy approximation algorithm for MDS. Starting from an empty set, if the current subset of vertices is not the dominating set, a newvertex which has the most number of the adjacent vertices that are not adjacent to anyvertex in the current set will be added.

The greedy algorithm approximates MDS within*1+ln(s)* where *s*is thesize of the largest set.

**SENTENCE GRAPH GENERATION**

* Each nodeis a sentence in the document collection.
* Sentences as vectors based on tf-idf, and then obtain the cosinesimilarity for each pair of sentences.
* If thesimilarity between a pair of sentences si andsj is above a given threshold λ, then there isan edge betweensi andsj.

**GENERATING MDS**

Starting from empty set, if the current subset of vertices is not in DS, a new vertex which highest number of adjacent vertex that are not adjacent to any vertex in the current set will be added.

*v\* = argmaxv s(v)*

Algorithm for generic summarization

Input: Sentence graph(G); Max length of summary(W)

Output: Min Dominant Set(S)

S=Ø

T=Ø

While L(S )< W and V(G) != S do

For v ϵ V(G)-S do

s(v) = |{ADJ(v)-T}|

v\* = argmaxv s(v)

S = S U {v\*}

T = t U ADJ(v\*)

**GENERATING MDWS for Query-Focused Summarization**

The queryfocused summarization can be modeled as

D\* =argminD⊆GΣs∈D d(s, q)

s.t. Dis a dominating set of G

d(s, q) can be viewed as the weightof vertex in G

v\* = arg maxvw(v)/s(v)