

Manipal Academy of Higher Education

Impressions@MAHE

Manipal Institute of Technology, Manipal
Theses and Dissertations

MAHE Student Work

Summer 7-1-2020

Semantic Similarity Based Abstractive Text Summarization and Evaluation Metric

Abishek .

Follow this and additional works at: <https://impressions.manipal.edu/mit>



Part of the [Computer Sciences Commons](#)

**Semantic Similarity Based Abstractive Text
Summarization and Evaluation Metric**

A postgraduate report submitted

to

MANIPAL ACADEMY OF HIGHER EDUCATION

For Partial Fulfillment of the Requirement for the

Award of the Degree

of

Master of Technology

in

Software Engineering

by

Abishek

Reg. No. 180928008

Under the guidance of

Dr. Sanjay Singh

Professor

Department of Information and Communication Technology

Manipal Institute of Technology, Manipal-576104



MANIPAL INSTITUTE OF TECHNOLOGY

MANIPAL

(A constituent unit of MAHE, Manipal)

JULY 2020

ABSTRACT

Text summarization is the process of producing a short and articulate summary by preserving vital information and meaning. Text summarization reduces the reading time and accelerates the process of studying the information. It is an exciting and challenging domain in Natural Language Processing (NLP) and has proliferated in recent years. Abstractive text summarization is a type of text summarization which generates human way of summaries, which are difficult.

The semantic similarity-based content selector module is a graph-based module that extracts the essential sentences from the source document. The extracted sentences are given to the abstractive summarization model. To evaluate the summaries we introduce a new semantic similarity-based evaluation metric, which compares how similar the reference and the system-generated summaries are, by calculating the cosine distance between the sentence embeddings of those summaries.

By visual observation, it is found that the quality of the summaries generated by the semantic similarity-based content selector module combined with abstractive summarization model is good. On evaluating the summaries generated by the State-of-the-art abstractive model using the ROUGE and the proposed metric, we found that the semantic similarity-based evaluation metric provides higher scores for the abstractive summaries by considering the novel's semantics words or sentences.

The semantic similarity-based content selector module, when combined with the abstractive summarization model, provides the quality abstractive summaries. The semantic similarity-based evaluation metric addresses the limitation of the ROUGE metric and shows that it is on par with human judgment.