How can Natural Language Processing summarizes educational videos?

Basile Masson

MSc Data Analytics and Artificial Intelligence

Prof. Cédric Verbeeck



EDHEC Business School does not express approval or disapproval concerning the opinions given in this paper which are the sole responsibility of the author.

I certify that the master project being submitted for examination is my own research, the data and results presented are genuine and actually obtained by me during the conduct of the research and that I have properly acknowledged using parenthetical indications and a full bibliography all areas where I have drawn on the work, ideas and results of others and that the master project has not been presented to any other examination committee before and has not been published before.

Draft & ideas

First tests, most interesting keyword extraction.

I tried the PKE library from Florian Boudin. First results are interesting, however, it only computes the best keywords with the appropriate scoring.

@InProceedings{boudin:2016:COLINGDEMO,

author = {Boudin, Florian},

title = {pke: an open source python-based keyphrase extraction toolkit},

booktitle = {Proceedings of COLING 2016, the 26th International Conference on Computational Linguistics: System Demonstrations},

month = {December},

year = {2016},

address = {Osaka, Japan},

pages = {69--73},

url = {http://aclweb.org/anthology/C16-2015}

}

First I start with extraction where I only choose the best sentences. Interesting but not good enough.

Gensim

* Text Summarisation with Gensim (TextRank algorithm)-
* We use the summarization.summarizer from gensim.
* This summarising is based on ranks of text sentences using a variation of the TextRank algorithm.
* TextRank is a general purpose, graph based ranking algorithm for NLP.
* TextRank is an automatic summarisation technique.
* Graph-based ranking algorithms are a way for deciding the importance of a vertex within a graph, based on global information recursively drawn from the entire graph.

Good results and easy to implement

**Now, I want to try Deep Learning for NLP. Instead of an extractive method, this is a abstractive method.**

Transformers: State-of-the-art Natural Language Processing for Pytorch and TensorFlow 2.0.

Introduction to Text Summarization

Text Summarization is one of the many applications of the Natural Language Processing (NLP) field, which is the subset of Artificial Intelligence which deals with interactions between computers and human (natural) languages.