

FNS-Summarisation 2020 shared task: system description paper

Extractive Summarization System for Annual Reports

Abderrahim Ait Azzi

Fortia Financial Solutions

17 av. George V, 75008

Paris, France

abderrahim.aitazzi@fortia.fr

Juyeon Kang

Fortia Financial Solutions

17 av. George V, 75008

Paris, France

juyeon.kang@fortia.fr

Abstract

In this paper, we report on our experiments in building a summarization system for generating summaries from annual reports. We adopt an "extractive" summarization approach in our hybrid system combining neural networks and rules-based algorithms with the expectation that such a system may capture key sentences or paragraphs from the data. A rules-based TOC (Table Of Contents) extraction and a binary classifier of narrative section titles are main components of our system allowing to identify narrative sections and best candidates for extracting final summaries. As result, we propose one to three summaries per document according to the classification score of narrative section titles.

1 Introduction

The Financial Narrative Summarization (FNS) aims to generate textual summaries from annual reports where financial overview of a company over a year stated. In this FNS shared task 2020 (El-Haj et al., 2020), we focus on "narrative" parts of those documents, all textual parts except tables, figures and diagrams. The provided dataset contains the narrative summaries as the gold standard and the expected summaries should be formed only from the narrative parts of the documents. These documents often contain more than 100 pages of texts, tables, figures, diagrams, etc. Summarizing an annual report of a company into a short text allows to get insights on annual financial status of the company.

The summarization approaches can be extractive or abstractive not only according to the techniques we adopt but also depending on the dataset on which we build the system. Our experiments are designed based on the extractive approach where we develop a pipeline in order to extract the most informative sections. We suppose that each narrative section is already a summary as the author of annual reports aim to resume key financial overview in each section. This is why we adopt the extractive summarization approach and the results of our system largely depend on the narrative sections prediction.

First, we create a binary classification dataset and implement Machine Learning algorithms to identify the narrative sections with its sections titles, then analyze which sections contain key information on annual financial status. Second, we select the summaries which capture the most informative parts of texts from the previously selected sections.

For this shared task, we experiment hybrid methods combining neural networks and rules-based systems in our summaries extraction algorithms, according to the results of the narrative sections prediction, a document can be assigned one to three summaries as described in the section 4. In the section 5, we shows that our system obtains encouraging ROUGE scores and discuss on the major errors that we could observe in our final summaries. We conclude the paper by proposing some future works.

2 Related work

Text summarization is one of the most challenging tasks in NLP. In general, two methods are used to handle this task, extractive and abstractive methods.

Extractive summarization methods produce summaries by selecting the most relevant sentences directly from the source text. Many approaches proposed for this type of summarization, some studies address the problem as a sentence-level sequence labeling task where each label indicates if the sentence

should be included in the summary or not, using either a CRF model (Nguyen MT., 2017) or using attention RNN-based Seq2seq Models as in (Cheng and Lapata, 2016; Nallapati et al., 2016). These models require generally a large and high-quality training sets. There are also some unsupervised methods in the extractive approach using statistical methods as a simple tf-idf (Saggion et al., 2016) or graph-based approaches as (Barrios et al., 2016) that represent text as a network linking sentences and use graph-based ranking methods to generate a summary.

Abstractive summarization, on the other hand, is a technique in which the summary is generated by generating novel sentences. This method requires then a deeper interpretation and understanding of the text and a text generation system. Most researches are based on a sequence to sequence attention based models and more recently transformers to handle this challenge (Zhang et al., 2020).

3 Data

The financial annual reports are publicly available documents on which firms publish a year-end summary of their operations and financial conditions. An annual report covers a broad scope of contents in different sections. The main data sets of the FNS shared task (see Table 1) have been taken from the UK firms annual reports, and previously converted into text file format.

Dataset	Train	Valid	Test
Annual report full text	3000	363	500
Gold Summaries	9873	1250	1673 (not released)

Table 1: Dataset.

The key narrative sections we have identified from the given train set are *Financial and Operational Highlights*, *Chief Executive’s Review*, *Financial Review*, etc., it means that those sections typically contain summarized financial issues. As those documents include a lot of statistical informations over a year, tables, graphs, diagrams and figures are often used for key informations along with textual statements, called *narrative* sections. We only focus on the *narrative* sections for building our summarization system.

4 System overview

As shown in Figure 1, We first identify narrative sections from the reports, then extract summaries from those sections. To do that, we extract the TOC from each document and train a binary classification model with Keras using the list of narrative sections found in the gold summaries as train set.

4.1 TOC extraction

In this first step, we design a rules-based algorithm extracting the table of contents (TOC) from a document. For that purpose, we segment a document into sentences, then each sentence is matched with a rules set in order to verify if it is a title of the TOC. We consider a sentence as title if it satisfies predefined conditions, some examples are given as follows : 1) a sentence starts with a digit, 2) it does not contain any of the predefined noisy words and special characters and 3) the word length in a sentence is less than 10. We applied this approach to all the documents in the training set.

4.2 Narrative sections prediction

The second step is to identify from the titles of the extracted TOC, the ones that correspond to narrative sections. For this purpose, we concatenated all the TOCs extracted from the training set and built a binary classification dataset with the labels: 1 if the title is narrative (a title is considered as narrative if it is seen in the gold summary), 0 otherwise. The labels attribution was not obvious, because some titles can be narrative in reports and non narrative in others. Therefore, we use the most frequently occurring label of each unique title as its ground truth label. Once we prepared the data set, we designed a binary classifier to predict if a section is narrative or non narrative. We implemented a Convolutional Neural

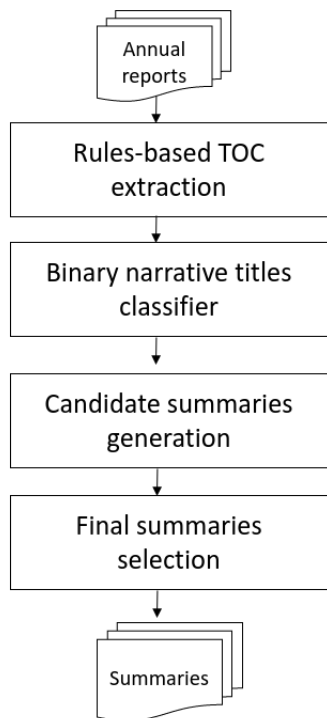


Figure 1: System overview

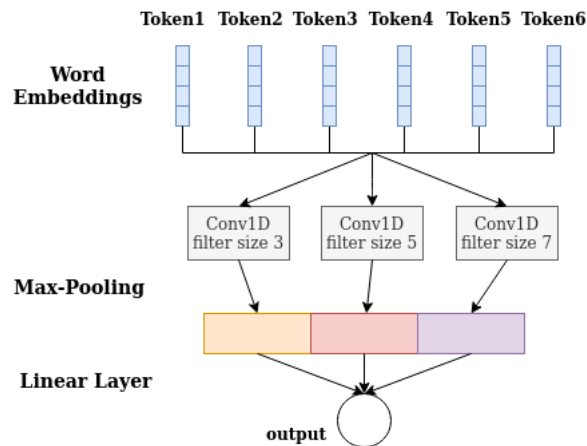


Figure 2: Binary classifier architecture

Network (CNN) classifier using Keras. Figure 2 (Mansar and Ferradans, 2018) shows the architecture of the model that contains: a word embeddings layer to give a dense representation of each token of a title. This representation passes through a convolutional network and pooling layer to encode the title. Finally, a linear layer with a sigmoid activation to predict the output. We split the dataset into a train set (80 %) and test set (20%). For the TOCs given in the test set, the classifier predicts if a title is a narrative section title or not with a score for each title and results the titles of narrative sections for each document. The trained model reaches a classification accuracy of around 70% on the test set.

4.3 Candidate summaries extraction

We identified the titles of narrative sections from the test set using our narrative sections titles classifier. For a report, the classification results show that we have in average 5~6 narrative sections knowing that the average sections number in a report is 29.5. We now generate "candidate" summaries by the way that we first select the top 3 predictions having the top classification scores among the narrative section titles of a report. Then, for each narrative title, we search of its first occurrence in the report in such a way that it is not a part of the TOC but a title of a section. Finally, we extract the first 1000 words of the report starting from this occurrence title. The extracted content of each of the top 3 titles corresponds to the system 1, 2 and 3, respectively.

4.4 Final summaries generation

In this last step, we focus on post-processing the "candidate" summaries extracted in section 4.3. The post-processing is mainly to check if the candidate summaries are not overlapping. This overlapping can occur due to the strategy of taking the first 1000 words starting from the narrative title. Therefore, for a candidate summary, if the range of the 1000 words contains contents from the next narrative section (hence the next candidate summary), we delete the overlapped contents from the first summary.

5 Results and Discussion

At the time of the deadline, the best summaries we had were the summaries gathered from the sections with the top 3 classification score, which are corresponding to the systems 1, 2 and 3, respectively. We also randomly selected all the final summaries of 100 reports and read the summaries focusing on the contents of each summary if it provides firm's financial overview of the business over the year. This defined which system's summary we should submit, whose results (as provided by the organizers) are reported in Table 2.

Metrics	System 1	System 2	System 3
R.L-R/R.L-P/ R.L-F	0.40/0.40/ 0.38	0.37/0.34/0.33	0.34/0.34/0.33
R.1-R/R.1-P/ R.1-F	0.43/0.43/ 0.41	0.39/0.41/0.38	0.37/0.37/0.35
R.2-R/R.2-P/ R.2-F	0.30/0.28/ 0.27	0.25/0.26/0.24	0.21/0.22/0.20
R.SU4-R/R.SU4-P/ R.SU4-F	0.34/0.33/ 0.32	0.30/0.31/0.29	0.26/0.27/0.25

Table 2: Results measured by the organizers for the test set.

The multiple variants of ROUGE measures, ROUGE-L, ROUGE-1, ROUGE-2 and ROUGE-SU4 are used for automatic evaluation. ROUGE (Lin, 2004) compares any summary to any other (normally human generated) summary. ROUGE -1 and ROUGE-2 are based on unigrams and bigrams, respectively, and ROUGE-SU4 uses bigrams with a maximum skip distance of 4 between bigrams.

With the information given in the evaluation results, we carried out an error analysis on the final summaries that we have submitted. We sum up the main elements which can be improved in the future works. First, noisy texts can be cleaned. We observed that many of the final summaries still contain pages number (see Figure 3) and tables (see Figure 4). Tables contain statistical informations which are not considered as narrative statement.

p6	
Strategy and priorities	
Our priorities and how	
we plan to deliver them	
p10	
Our KPIs	
How we measure our	
progress and delivery	
p12	
The environment we operate in	
Capitalising on opportunities	Strategic Report Corporate Governance Financial Statements Additional Information
The trends and factors impacting	National Express Group PLC Annual Report 2017 3 Strategic report roadmap 0.0
development of our industry	0.5
p4	1.0
Principal risks and uncertainties	1.5
The risks faced by National Express Group	2.0
and how we manage them	2.5
p38	3.0
Environmental, social and	3.5
employee issues	4.0
How we work with our principal stakeholders	UK Spain US Germany Morocco
p30	%
Our performance in 2017	2011 - 16 average 2017 - 19 average
Group Chief Executive's review	0
An analysis of our strategy, performance	1,000
and results in 2017	2,000
p14	3,000
Group Finance Director's review	4,000
A review of the key factors impacting	5,000
our financial performance	6,000
p19	2015 2030 1990
	Population (thousand)

Figure 3: Noisy text: pages numbers

Figure 4: Noisy texts: tables

Second, some titles predicted as narrative section are false positives. They are often mentioned in sections contents, for example, a narrative section title, *Chairman's statement* can be seen as a part of a sentence like *Chairman's statement on pages 2 and 3*. In both cases, *Chairman's statement* is considered as positive title and its content is extracted as a part of a summary. We also observe that a section title can be seen as narrative in a report and non narrative in other report. This needs taking into account contextual information in building the classifier. The other issue related to narrative titles prediction is seen in some section titles which are too general like *Introduction* and *At a glance*. Those sections are normally narrative sections but rarely give an interesting overview on the financial status and limited to mention general presentation of the firm's activities.

6 Future work

For future works, we plan to improve the TOC extraction algorithm by adding POS (Part-Of-Speech) tagging step before applying rules set to the dataset, this will help to implement new rules which discriminates some unusually used syntactic categories like verbs and modals in the TOC. TOC extraction results are used as the dataset for training our binary classifier of narrative titles, consequently, we expect that the best narrative sections prediction would be improved.

We also believe that giving a weight to some narrative sections titles may help selecting the more relevant summaries. An idea is to use the number of occurrences of each unique narrative title as a weight of its importance.

References

- F Barrios, F Lopez, L Argerich, and R Wachenchauser. 2016. Variations of the Similarity Function of TextRank for Automated Summarization.
- J. Cheng and M. Lapata. 2016. Neural summarization by extracting sentences and words. arXiv preprint arXiv:1603.07252.
- Mahmoud El-Haj, Ahmed AbuRa'ed, Nikiforos Pittaras, and George Giannakopoulos. 2020. The Financial Narrative Summarisation Shared Task (FNS 2020). In *The 1st Joint Workshop on Financial Narrative Processing and MultiLing Financial Summarisation (FNP-FNS 2020)*, Barcelona, Spain.
- Chin-Yew Lin. 2004. ROUGE: A Package for Automatic Evaluation of Summaries. In *Text Summarization Branches Out*, pages 74–81, Barcelona, Spain. Association for Computational Linguistics.
- Y Mansar and S Ferradans. 2018. Sentence Classification for Investment Rules Detection. In *Proceedings of the First Workshop on Economics and Natural Language*.
- Ramesh Nallapati, Feifei Zhai, and Bowen Zhou. 2016. SummaRuNNer: A Recurrent Neural Network based Sequence Model for Extractive Summarization of Documents. 1611.04230.
- Tran CX, Nguyen ML, Nguyen MT., Tran DV. 2017. Summarizing Web Documents Using Sequence Labeling with User-Generated Content and Third-Party Sources.
- Horacio Saggion, Thierry Poibeau, J. Piskorski, and R. Yangarber. 2016. Automatic Text Summarization: Past, Present and Future.
- Jingqing Zhang, Yao Zhao, Mohammad Saleh, and Peter J. Liu. 2020. PEGASUS: Pre-training with Extracted Gap-sentences for Abstractive Summarization.