Automated Text Summarization for Dialogues with Transformer Models

Specialisation Project 1 (VT1), Master of Science in Engineering, Zurich University of Applied Sciences

Loran Avci avci@zhaw.ch

January 31, 2021

Contents

Abstract

This thesis is dedicated to the automated summarisation of texts and dialogues and covers the implementation of different approaches. While text summarisation is already a well-known topic in the Natural Language Processing (NLP) domain, there are only a few approaches that specialise in summarising dialogues between several parties in a user-friendly way (e.g. to-do list that emerges from a meting transcript). It is shown that the classification of action items in dialogues is technically possible with the help of RoBERTa which achieves very good results. Another approach proposes the supervised sequence-to-sequence generation of a summary with BART. However, the available data which is needed to train both of these models, is not mature enough and too domain specific. If new corpora or datasets emerge that address these shortcomings, the approaches should be revisited in future work.

In another attempt, which is based on transcripts of political debates, the individual utterances are assigned to dialogue type categories. The longest 10% of the utterances are monologues which are abstractly summarised using a transformer model and on average still have 25% of their original length. The next 15% longer utterances are statements, which are also summarised using a transformer model but contain a constant length of 5-15 tokens. The remaining utterances are merged into discussion blocks of a parameterisable length. TF-IDF is then used to extract the most important key words from the discussion blocks. The extracted key words are then assigned to the individual utterances. This algorithm delivers results that are easy to interpret and provide a high degree of relevant information. The user is thus able to avoid 90% of the original transcript, thus saving a lot of time.

1 Introduction

Automated summarisation is generally defined as the process of computationally summarising a given set of data to produce a new set of data that represents the most important and/or relevant information from the original data. For texts, this specifically means that the summary contains the most important sentences, text fragments or words (Torres-Moreno, 2014). A text summary should, if possible, be coherent, concise and fluent. In general, there are two methods of automatic text summarisation. One approach is extractive-based summarization, in which the content is extracted from the original text that seems relevant for summarization. However, the content of the extracted data is not modified but remains in its original form. Another approach is abstractive-based summarisation. This is a summary that is similar to one written by a human. Here, the content of the original text is paraphrased, for example, to obtain a more compact summary. Abstractive-based approaches usually require large amounts of training data to train a model and are often more computationally intensive than extractive-based approaches (Kathri et.al., 2018).

1.1 Initial situation

The field of automated text summarisation using NLP and neural networks has gained momentum with the introduction of transformer models (Vaswani et. al., 2017). Summarising continuous text, such as newspaper articles, is no longer a major challenge and is already being done with good results (Lewis et. al., 2019). However, the literature research has shown that little knowledge is available for dialogues in which several parties speak, and only a few approaches have been developed, respectively. One possible approach, for example, is done by Sentence-Gated Modeling Optimized by Dialogue Acts (Goo and Chen 2018). One criticism of the existing approaches is that most of them use the AMI corpus for validating their models, which contains very uniform and domain-specific summaries. Thus, it is not clear to what extent these models can be generalised.

1.2 Objective

The motviation for this work comes from the application “Interscriber” (see Figure 1) from which audio recordings of interviews, meetings and discussions can be transcribed automatically (SpinningBytes AG, 2021). When looking at business meetings or political debates that can last several hours, it can be helpful for the user if not only a transcript but also a resulting summary is available. This summary should be configurable in terms of length and have different levels of detail. Furthermore, the summary should still show which speaker makes certain statements at what time. This work tries to develop and implement such an approach.

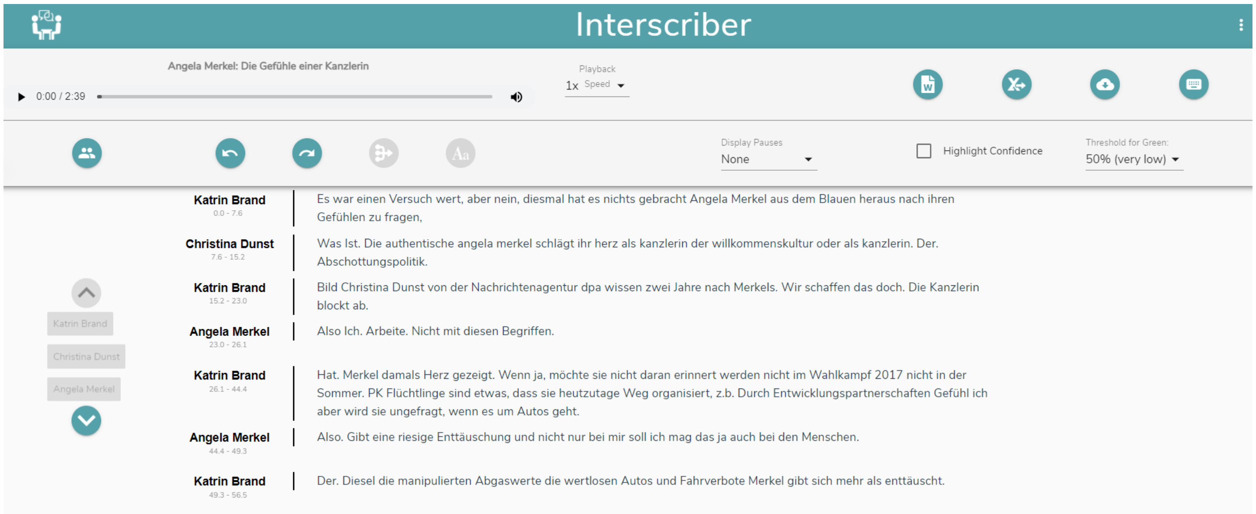

Figure 1: The figure shows the user interface of the Interscriber application for transcription of audio files (SpinningBytes AG, 2021).

2 Methods

This section explains which approaches were considered in the course of this work. It shows which methods are used and which data can be used. It also shows which literature the methods refer to and where the data can be found. It should be noted that first two approaches are shown which were not purposeful for this work under the given circumstances. Subsequently, the idea and implementation for the most promising approach are explained in more detail.

2.1 Action Item Classification

This approach is based on the assumption that statements containing an action item are relevant to a summary. Business meetings, for example, take place at regular intervals and the number of participants remains more or less the same. Discussions within such meetings usually contain statements that indicate what each participant should do between the current and the next meeting (Murray and Renals, 2008). Detecting and extracting such action items would come very close to summarising them in a business meeting.

An action item should fulfil certain characteristics. The content of an utterance should contain a concrete future action which would be noted in a to-do list, for example. Furthermore, this action should be explicitly assigned to a person who can carry it out, as well as their consent to it. Finally, there should be a timeframe when such an action should be carried out. It follows that an action item can be divided into the four subclasses: task, owner, timeframe and agreement (Purver et. al., 2006).

Such approaches have already led to promising results, for example in (Morgan et. al., 2006). Here the authors try to classify action items in multi-party audio meeting recordings. They use a maximum entropy model and compare the effect on performance for different selected features (e.g. semantic, syntactic and temporal features). The paper also points out that the data used, which corresponds to the ICSI corpus, is highly unbalanced and there is a low inter-annotator agreement (Janin et al., 2003).

To pursue such an approach, it would be necessary to have data in which statements containing action items are annotated as such. The literature search revealed that there have been attempts to classify action items in meetings using embeddings created by convolutional deep structured semantic models (Chen and Hakkani-Tür, 2016). The authors present a dataset in which action items are annotated by humans. The embeddings they create then serve as feautres which they classify using a Support Vector Machine. The results they obtained are promising, making the data potentially to a good foundation.

Criticisms of this approach include the work of McGregor and Tang, 2017, who go one step further and skip the transcription step, making it an Automatic Speech Recognition problem. In their paper, they try to detect action items in multi-party meetings that are converted into a task by a personal assistant such as "Siri" without giving a direct command to the personal assistant. The authors mention the complexity inherent in the nature of a meeting. Meetings are not passive actions that merely serve to collect data, as the statements that are made are not always clear and require interpretation from the participants (McGregor and Tang, 2017).

Furthermore, there are approaches that do not detect action items but rather utterances in which a decision is made. The methods used in these papers are no different from those used to identify action items. However, they can serve as additional datasets, as decisions are annotated in the data used (Deleris et. al., 2018).

The latter results have shown that neural approaches also have their justification. In a recently published paper, it is shown how the Bidirectional Encoder Representations from Transformers (BERT) model presented by Google can outperform the state of the art in various NLP tasks (Devlin et. al., 2019). The author of the paper also focuses on the area of multi-party meetings. By using BERT, the author of the paper is able to achieve promising results for the classification of action items (Sheshadri, 2019).

2.1.1 ICSI Corpus

The ICSI Corpus meeting corpus is a dataset in which natural meetings have been recorded at the International Computer Science Institute in Berkley. The corpus is hand-transcribed and contains various meta-data. The corpus is versatile and can be used for automatic speech recognition, information retrieval, etc (Janin et.al., 2003). In the paper of Chen and Hakkani-Tür, they present an extended dataset based on a subset of the ICSI Meeting Corpus. In it, action items are annotated which could be generated by an automated meeting assistant, for example. The data is created by marking utterances that contain potential actions or suggest future actions, for example, when a meeting participant is asked to search for a specific email or when a meeting date is set.

However, it should be noted that when annotating the data, the average agreement on whether a statement contains an action item or not is 64%. Ten different types of action items are defined in the dataset. The match for the type of action item is 100%. The different types of action items include the domains calendar, reminder, communication and others. The dataset contains 21000 statements, 318 of which are labelled as action items.

2.1.2 Bert for Text Classification

Consider we have 2 columns one which contains the utterance and another which contains the type of action item. In the first step we convert the action item into a binary variable since our goal is only to detect relevant information in a meeting transcript rather than identifying different types of action items. The data is then preprocessed, which means that the utterances are converted to lowercase letters, the punctuation is removed and the words are lemmatised. After the preprocessing, the data is stratified divided into training and test sets, with 20% of the data used for testing (In this section of the project we intentionally do not divide the data into training set, validation set and test set).

To classify the data adequately, a classification model is loaded from the library `simpletransformer`. The model type used corresponds to the model RoBERTa Base. This is a pretrained Transformer model based on a robustly optimised BERT Base model. The architecture of BERT Base contains 12 layers, 768 hidden dimensions and 110 million parameters (Devlin et. al., 2019). The Roberta model adds about 15 million parameters (Liu et. al., 2019).

The model achieves a Roc AUC of 0.8951 on the test set after 1 training epoch. In the test set, 61 action items were correctly classified, 16 action items were not recognised. Of the 4130 utterances that did not contain an action item, only 8 were incorrectly classified. The performance metrics can be found in table 1.

	precision	recall	f1-score	support
No Action Item	1.00	1.00	1.00	4130
Action Item	0.88	0.79	0.84	77
accuracy			0.99	4207
macro avg	0.94	0.90	0.92	4207
weighted avg	0.99	0.99	0.99	4207

Table 1: The table shows the performance values for the binary classification of action items using BERT.

While the performance measures were convincing, the evaluation of the results showed that the data including the labels were not up to the task of this work. The reason for this is that many utterances do not contain the desired information that one would expect for a summary of a dialogue transcript. Many utterances that contain an action item are ambiguous from both an objective and subjective point of view (see table 2). These findings show that it is technically possible to recognise action items in dialogues and that it can lead to a desired result when applied to meetings. Under the given circumstances, however, it is not possible to achieve a satisfactory result with the available data. Therefore, this approach will not be pursued further in this paper.

Utterance	Label
The - the meeting is July sixteenth through eighteenth .	Action Item
There's a median filtering and then there's a piece-wise linear fit, based on some criteria . I'm not sure.	Action Item
O_K. I can get the address, or if you know where you are, I can tell you how to get there.	No Action Item

Table 2: The table shows some utterances with their corresponding labels. Whether an utterance contains an action item or not can often be ambiguous.

2.2 Supervised Text Generation

In this experiment, it is assumed that a model can be trained to summarise texts, provided that an extractive or abstractive summary exists. The model should be able to generate new text from an input text, which in this case should be a summary of a predefined length. Furthermore, the model should then be generalised enough so that not only the meetings of the training data set can be summarised, but also any transcripts of dialogues with any number of speakers. In addition, the model should be able to produce useful summaries regardless of the domain of the input text.

2.2.1 AMI Corpus

The literature search has resulted in a dataset containing transcripts as well as extractive and abstractive summaries. This is the corpus of the Augmented Multi-party Interaction (AMI) project which contains over 100 hours of meetings. The corpus includes both, reallife meetings and role-played meetings that follow a specific scenario. The corpus also includes the transcripts as well as the associated speakers and both extractive and abstractive summaries for all meetings following a scenario (Mccowan et. al., 2005).

Source codes parsing the XML files into TXT files can be found on github (Sergio, 2019). The corpus contains the transcripts of 167 meetings. The meeting ID and speaker ID can be extracted from the file names. Of the transcripts, 137 have an extractive summary and 142 have an abstractive summary. Only the transcripts with both types of summary were used for the implementation of this approach.

2.2.2 BART for Text Generation

The data is aggregated in such a way that the statements of the different speakers are arranged one after the other. In this way, the complete input text can be assigned to each meeting ID.

A problem that arises from this is that the chronology of the statements is lost because the data does not contain timestamps. This means that the statements within a speaker run chronologically, but the chronology between the different speakers is lost. This limitation is ignored for the moment. After the input texts are aggregated accordingly, each input can be assigned a target text that corresponds to an extractive or abstractive summary.

A classical preprocessing is not carried out afterwards. The reason for this is that a model based on BERT is used (Lewis et. al., 2019). Since the BERT model uses wordpiece embeddings, it is not necessary to perform preprocessing of the text such as tokenization (Devlin et. al., 2019). Wordpiece embeddings divide words into a limited set of common sub-words. This method combines the flexibility of character delimited models with the efficiency of word delimited models (Wu et. al., 2016). To generate a target text from the input text, a seq2seqmodel is loaded from the `simpletransformer` library. The model used for the summaries is the BART-large-cnn. This is a pre-trained model that combines bidirectional and autoregressive transformer architecture which was finetuned on summarising CNN News articles. BART is trained by processing the text using a noise function and then learning a model to reconstruct the original text. The architecture of BART is similar to that of BERT - the base model contains 6 layers in both the encoder and the decoder, while the large model contains 12 layers each. BART also does not use feed forward networks for word prediction. BART contains about 10% more parameters than the corresponding BERT model (Lewis et. al., 2019).

One disadvantage of the model is that it cannot process more than 512 characters at a time, which leads to the majority of a text being truncated. To prevent this, the input text is first broken down into individual sentences. The sentences are then divided into nests with a maximum length of 512 characters. A summary can then be created for each of these nests. Finally, the individual summaries of the nests can be merged to function as the target text of the input text. In order for the model to be able to create the summary of the meetings, the pre-trained model is fine-tuned. For this purpose, the data is divided into a train and a test set, with the test set containing 20% of the data. In this section of the project we intentionally do not divide the data into training set, validation set and test set. The model is thus passed the input texts and the target texts in order to fine-tune it. Subsequently, only the input texts are passed to the model in order to predict the target texts.

A first look at the predicted target text, in other words the summaries of the input texts, shows that the model is strongly over-fitted to the training data. This is due to the fact that all input texts correspond to a pre-played scenario which always contains the same procedure. Furthermore, the content of the scenario meeting is similar each time and does not deviate from a specific domain. The results are then not considered or interpreted any further. No performance measures are calculated, as it can be assumed that an over-fitted model can only produce useful predictions or summaries with regard to the training data. This approach has shown that given the AMI corpus data, it is not possible to train a supervised seq2seq model that can be generalised to other dialogues. Therefore, this approach is no longer pursued.

2.3 Unsupervised Text Generation and Key Word Extraction

This approach is the one that ultimately leads to the goal of this work. It is based on a concept which is derived independently in this work. The approach is based on the author's assumption that utterances of different lenghts contain different degrees of information. All utterances are assigned to one of three categories depending on the number of tokens. For short utterances, only the key words are extracted, while for longer utterances, abstract summaries are created using a transformer model.

The literature search revealed that segmenting a discourse into different types led to promising results (Bokai et. al., 2015). The paper mentions that the length of an utterance is the best feature to indicate whether a sentence should be summarised. The authors also segment their dialogues into monologues, in which a single speaker dominates the discourse, and discussions, in which multiple speakers are involved in the discourse.

The transformer model is a pre-trained T5 model which is not fine-tuned any more. This is to prevent the model from over-fitting itself to a certain type of dialogue. The threshold value for the categories depends on the distribution of the token length of all utterances. This guarantees that the length of the summary is always the same in real terms. Nevertheless, the length of the summaries as well as the number of keywords can always be set by the user through parametrisation. In figure 2 a visual concept is shown, which shows the hierarchical summary as well as the approaches for different text lengths. In a first step, the dialogue is divided into different areas in which the individual utterances can each be assigned to a theme. Then, on a second level, the summary for long utterances is displayed as well as the key words for the individual utterances.

2.3.1 US Election 2020 Data

To implement this approach, data is required which has utterances that are annotated to a speaker in correct chronological order. For the implementation of this approach, data provided by the data science community Kaggle was used (Kaggle, 2021). The data are transcripts of political debates from the 2020 US Presidential Election, featuring Republican President Donald Trump and his Democratic opponent Joe Biden. Also debating are Republican Vice President Mike Pence and his Democratic opponent Senator Kamala Harris. The debates will each be moderated by a male or female moderator. Thus, the transcripts each contain 3 parties, namely a Democrat, a Republican and a moderator. The dataset provides the debates both as audio recordings and as unstructured texts or cleaned csv.

In the analysis of the data, the data set us_election_2020_vice_presidential_debate.csv (VP-debate) and us_election_2020_2nd_presidential_debate.csv (P-debate) are taken into account. The dataset VP-debate contains 327 observations and three attributes. The P-debate dataset contains 512 observations and three attributes. The three attributes are as follows: 'Speaker', who is assigned the utterance that is currently being pronounced, 'minute', at what time after the start of the debate an utterance was made and 'text', transcribed text of an utterance. The attribute is not continuous because it resets to zero at certain times, usually after about 30 minutes. The transcripts of the audio files finally contain about 90 minutes for the VP-debate and about 98 minutes for the P-debate.

The distribution of utterances differs between the two data sets in the following ways. If we consider each observation as an utterance, it contains a median of 75 characters (mean = 259 characters) for the VP-debate dataset, which combine to form a median of 14 tokens (mean = 45 tokens) and a median of two sentences (mean = 3 sentences). For the data set P-debate, the median is 66 characters (arithmetic mean = 195 characters), the median is 12 tokens (arithmetic mean = 35 tokens) and the median is 2 sentences (arithmetic mean = 3 sentences). The distribution for the number of characters, tokens and sentences is unimodal and right-skewed. The histograms can be seen in Figure 3. The remaining evaluations and results refer exclusively to the VP-debate dataset.

If we look at the relative distribution in the VP-debate of utterances and the length of the utterances, we find interesting characteristics. The relative proportion of spoken tokens (individual words) for speaker 'Susan Page' is around 21%, for speaker 'Kamala Harris' around 38% and for speaker 'Mike Pence' around 40%. The relative proportion of spoken utterances (observations) is 36% for speaker 'Susan Page', 35% for speaker 'Mike Pence' and 29% for speaker 'Kamala Harris' (see figure 4). It can be concluded from this that Speaker 'Susan Page' in particular makes more utterances, which has a relatively smaller number of tokens. In contrast, speaker 'Kamala Harris' makes fewer utterances than the other parties, but they have a relatively higher number of tokens.

Figure 2: The figure shows a visual concept for the automated summary of dialogs. In a first step, the transcript is divided into thematic areas. In a second step, the summaries for specific utterances are displayed.

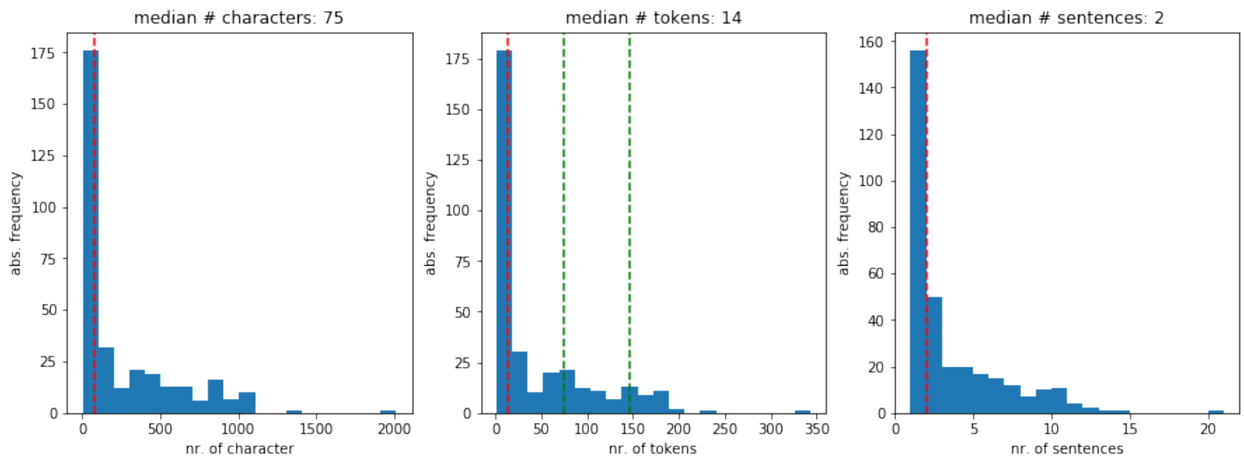

Figure 3: The figure shows the distribution of the number of characters (left), number of tokens (middle) and number of sentences (right). The red dashed line represents the median of the observations. The green dashed line represents the 75% and the 90% quantile, respectively.

2.3.2 Distribution-based Attribute Determination

To implement this approach, another attribute is created which uses the number of tokens per observation. Each observation is assigned one of three dialogue types. Observations containing a "small" number of tokens are assigned the dialogue type <discussion>, observations containing a "medium" number of tokens receive a <statement> token and observations with a high number of tokens receive a <monologue> token. The thresholds for small, medium and long numbers of tokens are determined using the distribution of the data. Thus, they always adapt to the data, which leads to consistent results. More precisely, two quantiles are determined that separate <monologue> from <statement> and <statement> from <discussion>. In this work, the 90% quantile for a <monologue> and the 75% quantile for

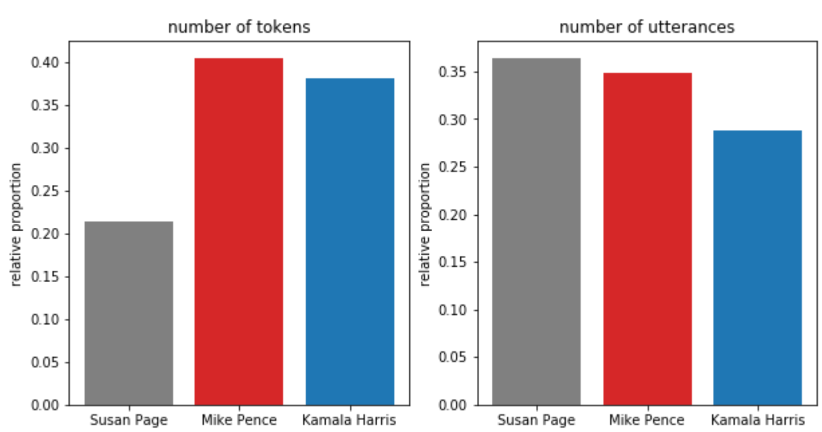

Figure 4: The figure shows the relative proportion of spoken tokens (left) and spoken utterances (right).

a <statement> token are chosen as thresholds (see figure 3). The chosen thresholds are, however, parameterisable and can be adjusted according to need. In general, it can be said that lower thresholds lead to more detailed summaries and higher thresholds to more compact summaries. For the VP-debate dataset, the threshold for the dialogue type <monologue> is at least 147 tokens and for the dialogue type <statement> at least 74 tokens. From this we can conclude that in the case of the VP-debate, the 75% of the utterances with the smallest number of tokens receive the dialogue type <discussion> (245 observations), the 15% of the next longest number of tokens receive the dialogue type <statement> (49 observations) and the 10% of the utterances with the longest tokens receive the dialogue type <monologue> (33 observations). The motivation behind this distinction lies in the assumption that longer statements contain more information. This leads to the approach that the different types of dialogue must also be summarised differently.

2.3.3 T5 for Text Generation

Text-to-Text Transfer Transformer (T5) is a transformer model that is very similar in architecture to BERT (Vaswani et. al., 2017). The model is pre-trained on the Colossal Clean Crawled Corpus and can perform different NLP tasks such as translation, question answering, summarisation and classification. The model used here uses the properties of a pre-trained model and transfers its knowledge to the data used here without fine tuning the model. The T5 Small model, consist of 12 blocks, in both the encoder and decoder. The feed-forward networks in each block consist of a dense layer with an output dimensionality of 2048 followed by a ReLU nonlinearity and another dense layer. All attention mechanisms have 8 heads. All other sub-layers and embeddings have a dimensionality of 512 (Raffael et. al., 2020). The T5 model is available in different sizes with the base model having over 220 million parameters. The model used in this work, T5 Small, has 60 million parameters. Furthermore, there are models with 770 million (T5 Large), 3 billion (T5 3B) and 11 billion (T5 11B) parameters. The pre-trained models are used through the `transformers` python package.

In the model specifically the values for the argument `num_beams` are increased to 32 which controls the number of beams for the beam search algorithm. Furthermore the argument `no_repeat_ngram_size` is increased to 3. This allows to control how often a n-gram in this case a tri-gram can occur in a summary.

The concept is that utterances with the dialogue type <monologue>, which make up only 10% of the data, receive an abstractive summary generated by a transformer model. The length of this summary is at least 20% and at most 40% of the original length. By setting limits, it can be guaranteed that automatic summaries do not become too lengthy. However, the chosen limits can be parameterised as desired.

A similar approach is used for utterances with the dialogue type <statement> in the VP-debate dataset, which have a length of 74 to 146 tokens. Abstractive summaries are created using a transformer model, but the length of the summary is not based on the original length, but a fixed length of 5-15 words. This still corresponds to a range of 5% of the lower threshold and 10% of the upper threshold. The limits for the number of words can also be adapted to specific needs if necessary.

2.3.4 TF-IDF for Key Word Extraction

For the utterances with dialogue type <discussion>, which makes up the bulk of the data, abstractive summaries have been shown to make little sense. Many utterances contain little information or consist only of stop words. Therefore, no transformer model is used here, but a key word extraction with the help of TF-IDF is carried out. Where TFIDF is defined by:

$$tf.idf(t, D) = tf(t, D) \cdot idf(t)$$

with

$$tf(t, D) = \frac{\#(t, D)}{max_{t' \in D} \#(t', D)}$$

giving us the Frequency of a term t in a document D and

$$idf(t) = log \frac{N}{\sum_{D:t \in D} 1}$$

where N is the Number of douments and $\sum_{D:t \in D} 1$ is the number of documents which contain the term t (Manning et.al., 2008).

First, utterances are concatenated in their chronological order with the constraint that concatenated utterances contain a maximum number of tokens that does not exceed the threshold for the dialogue type <statement> (e.g. 74 tokens).
The data is then processed and cleaned by performing various pre-processing steps typical for text data, such as removing stopwords. Furthermore, only those tokens are retained in which the part of speech refers to one of the following: NN Noun (singular or mass), NNS (Noun, plural), NNP (Proper noun, singular), NNPS (Proper noun, plural) (Marcus et.al., 1993). Through this step, mainly nouns are extracted as key words. Each of the concatenated discussion blocks thus corresponds to a document containing the tokens to be analysed. The list of documents can then be passed to a TF-IDF vectorizer which generates the features for a TF-IDF transformer. Finally, the three words that receive the highest score are declared as key words and assigned to their respective discussion block. The number of keywords is freely selectable and can be parameterised. In a final step, the key words are assigned to the individual utterances which form the corresponding discussion block. For both the utterances with dialogue type <monologue> and those with dialogue type <statement>, key word extraction was performed using the same procedure as for the utterances with dialogue type <discussion>. However, the individual utterances were not chained together in blocks, but were each considered as a separate documents.

2.3.5 Results

The finished output comes in the form of a formatted Excel spreadsheet which allows the user to get a quick overview of the dialogue. The resulting table consists of six columns containing: the name of the speaker, the time, the dialogue type, the key words, the summary (if the dialogue type is <monologue> or <statement>) and the full text of an utterance. By default, a filter is already configured which, when the Excel file is opened, displays only the lines for which the dialogue type is <monologue> (see figure 5). The column Speaker and dialogue type get an automatic colour coding to create a better overview. With the thresholds used in this work, the relative size of this first view will always remain the same. When opening the Excel file, the user will see 10% of the utterances from the full transcript in a compact form (e.g. transcript: 300 observations, output file: 30 observations). By configuring the filter in the output file, more detailed summaries can be displayed if required. For example, the dialogue type can be extended by the type <statement>. This has the consequence that all observations with this attribute are now displayed. The resulting output file is now expanded by 15% of the original transcript size (see figure 6). However, the summary does not become proportionally longer by 15% because, as already described, utterances with the dialogue type <statement> are summarised with a maximum of 15 words, while utterances with the dialogue type <monologue> have a summary length of at least 30 words (with the thresholds used in this work).

The most detailed version of the summary is obtained by configuring the filter on the dialogue type column to show all utterances with dialogue type <discussion> (see figure 7). The dialogue type is composed of a <discussion> token and a number preceded by a # character. The number indicates which utterances have been merged to perform the key word extraction. This enables the user to better interpret the output of the key word extraction and makes it easier for him to assign the key words to individual utterances. It is important to note that the summary column for utterances with dialogue type <discussion> does not contain an abstractive summary but only an token. As the name suggests, this can be ignored. It is possible to replace the token with any other token or with an empty string.

Furthermore, it is possible for the user to take a closer look at individual discussions, if the topic interests him or her, by means of keyword extraction. For example, the key words # tax # cut # break appear in the transcript of the VP-debate under the dialogue type <discussion #22>. The user can now filter out this discussion and read it in full or, for example, mark it for other users (see table 3). Alternatively, the output can also be viewed as html.

2.3.6 Summary Lengths

The evaluation of the summary lenghts refers exclusively to the data set VP-debate. To get an idea of the length of the summary, the lengths of the individual summaries are divided by their reference length respectively the length of the original text. For simplicity, utterances with dialogue type <discussion> are ignored here, as there is no summary for such utterances, only key words. The length of the summary can be calculated and interpreted on different levels. For utterances with the dialogue type <monologue>, the average length of the summary

speaker	time	dialogue type	key words	summary	text
Susan Page	00:00	<monologue>	# debate # audience # candidates	it is my honor to mo	Good evening. From
Kamala Harris	03:21	<monologue>	# ledger # plan # knew	president's plan is al	Can you imagine if yo
Mike Pence	04:55	<monologue>	# world # time # decision	from the very first d	Susan, thank you. An
Mike Pence	05:57	<monologue>	# lives # look # plan	more than 115 millic	And I believe it saved
Mike Pence	10:11	<monologue>	# people # americans # states	president Trump anc	Well, the American p
Susan Page	13:24	<monologue>	# concerns # vice # president	the president's diagi	Vice President Pence
Mike Pence	14:28	<monologue>	# vaccine # reality # failure	in unheard of time, i	Well, thank you, but
Mike Pence	15:29	<monologue>	# flu # year # wrong	if the flu had been a	It was 2009. The Swir
Kamala Harris	16:53	<monologue>	# dignity # age # joe	the day I got the call	Let me tell you first c
Kamala Harris	17:58	<monologue>	# woman # department # states	i was elected the firs	I was elected the firs
Kamala Harris	21:01	<monologue>	# 750 # decisions # debt	we now know Donal	Absolutely. And that
Susan Page	23:26	<monologue>	# jobs # growth # taxes	on friday, we learne	... which is about the
Kamala Harris	25:17	<monologue>	# infrastructure # ll # money	it's about upgrading	And through a plan tl
Mike Pence	27:29	<monologue>	# jobs # biden # trade	we've already addec	I mean, right after a t
Mike Pence	06:57	<monologue>	# auto # workers # jobs	senator Kamala Harr	Thank you, Susan. W
Mike Pence	08:06	<monologue>	# china # travel # coronavirus	president Trump ma	So now with regard t
Kamala Harris	10:10	<monologue>	# trade # war # disease	a reputable research	There was a team of
Mike Pence	14:09	<monologue>	# isis # ve # president	we stood strong witl	Thank you. Well, Pre
Kamala Harris	16:49	<monologue>	# strike # feels # hero	what happened to y	First of all, to the Mu

Figure 5: The figure shows the resulting summary filtered by the default dialogue type <monologue>.

speaker	time	dialogue type	key words	summary	text
Susan Page	00:00	<monologue>	# debate # audience # candidates	it is my honor to mo	Good evening. From
Susan Page	01:21	<statement>	# administration # week # states	39 states have had n	These are tumultuou
Kamala Harris	02:13	<statement>	# people # workers # affect	210,000 dead people	Thank you, Susan. W
Kamala Harris	03:21	<monologue>	# ledger # plan # knew	president's plan is al	Can you imagine if y
Susan Page	04:20	<statement>	# death # population # toll	more than 210,000 a	Thank you, Senator H
Mike Pence	04:55	<monologue>	# world # time # decision	from the very first d	Susan, thank you. An
Mike Pence	05:57	<monologue>	# lives # look # plan	more than 115 millic	And I believe it save
Kamala Harris	07:10	<statement>	# vice # president # hasn	the vice president kı	Absolutely. Whateve
Mike Pence	08:45	<statement>	# dr # reality # fauci	the reality is Dr. Fau	But the reality... If I n
Mike Pence	10:11	<monologue>	# people # americans # states	president Trump anc	Well, the American p
Mike Pence	11:07	<statement>	# government # mandates # people	the difference here	If I may say, that Ros
Kamala Harris	11:58	<statement>	# people # administration # food	this administration s	Speaking of those th
Susan Page	13:24	<monologue>	# concerns # vice # president	the president's diagi	Vice President Pence
Mike Pence	14:28	<monologue>	# vaccine # reality # failure	in unheard of time, i	Well, thank you, but
Mike Pence	15:29	<monologue>	# flu # year # wrong	if the flu had been a	It was 2009. The Swir
Kamala Harris	16:53	<monologue>	# dignity # age # joe	the day I got the call	Let me tell you first c
Kamala Harris	17:58	<monologue>	# woman # department # states	i was elected the firs	I was elected the firs
Susan Page	18:58	<statement>	# information # president # turn	neither president, n	Thank you, Senator H
Mike Pence	19:26	<statement>	# prayers # concern # forth	the care of the presi	Well, Susan, thank y
Mike Pence	20:03	<statement>	# expressions # congratulate # nomination	the american people	... they will continue

Figure 6: The figure shows the resulting summary filtered by the dialogue type <monologue> and <statement>.

speaker	time	dialogue type	key words	summary	text
Susan Page	00:00	<monologue>	# debate # audience # candidates	it is my honor to mo	Good evening. From
Susan Page	01:21	<statement>	# administration # week # states	39 states have had n	These are tumultuou
Kamala Harris	02:13	<statement>	# people # workers # affect	210,000 dead people	Thank you, Susan. W
Kamala Harris	03:21	<monologue>	# ledger # plan # knew	president's plan is al	Can you imagine if yo
Susan Page	04:17	<discussion #0>	# thank # president # vice	<ignored>	Thank you, Senator H
Kamala Harris	04:18	<discussion #0>	# thank # president # vice	<ignored>	... right to reelection
Susan Page	04:20	<statement>	# death # population # toll	more than 210,000 a	Thank you, Senator H
Mike Pence	04:55	<monologue>	# world # time # decision	from the very first d	Susan, thank you. An
Mike Pence	05:57	<monologue>	# lives # look # plan	more than 115 millic	And I believe it save
Susan Page	06:59	<discussion #0>	# thank # president # vice	<ignored>	Thank you, Vice Pres
Mike Pence	06:59	<discussion #0>	# thank # president # vice	<ignored>	... of America first. A
Susan Page	07:03	<discussion #0>	# thank # president # vice	<ignored>	Thank you, Vice Pres
Mike Pence	07:04	<discussion #0>	# thank # president # vice	<ignored>	... of the sacrifices th
Susan Page	07:07	<discussion #0>	# thank # president # vice	<ignored>	Thank you, Vice Pres
Kamala Harris	07:10	<statement>	# vice # president # hasn	the vice president kı	Absolutely. Whateve
Susan Page	07:52	<discussion #0>	# thank # president # vice	<ignored>	Well, let's [crosstalk
Kamala Harris	07:53	<discussion #0>	# thank # president # vice	<ignored>	No. But Susan, this is
Mike Pence	07:55	<discussion #1>	# mr # seconds # ll	<ignored>	Susan, I have to weig
Kamala Harris	07:56	<discussion #1>	# mr # seconds # ll	<ignored>	Mr. Vice President, I'
Mike Pence	07:57	<discussion #1>	# mr # seconds # ll	<ignored>	I have to weigh in.

Figure 7: The figure shows the resulting unfiltered summary. The user is thus able to follow discussions of interest with the help of the key words.

Speaker	Time	Key Words	Utterance
Mike Pence	29:14	# tax # cuts # break	The important [inaudible] is you said the truth. Joe Biden has said it twice in the debate last week that he's going to repeal the Trump tax cuts. That was tax cuts that gave the average working family $2,000 in a tax break every single year-
Kamala Harris	29:28		That is-
Mike Pence	29:28		Senator, that's the math.
Kamala Harris	29:30		That is absolutely not true, [crosstalk 00:00:29:32].
Mike Pence	29:32		Is he the only going to repeal part of the Trump tax cuts?

Table 3: Example of <discussion #22> with key words # tax # cuts # break.

is 25% of the original text. Since our parameters allow a maximum length of 40% of the original text, this may indicate that our model prefers more compact summaries for longer input texts. For utterances with the dialogue type <statement>, the average length of the summary can be 11% of the original text. The average length of the summaries of utterances with the dialogue type <monologue> and <statement> together is 17% of the length of the original texts. Because only the key words are extracted for utterances with the dialogue type and no summary is generated, the average length of the summary here is 0% of the length of the original texts. Averaged over the utterances for all dialogue types, the average summary length is 4% of the length of the original texts. Since the concept is based on the fact that the user only has to read the summaries of the utterances with dialogue type <monologue> to get an overview of the transcript, the aggregated length of these summaries is also compared with the total length of the transcript. All summaries with dialogue type <monologue> contain 10% of the total length of the transcript (see Table 4).

<table>
<tr><th>dialogue type</th><th colspan="3">avg. summary length</th></tr>
<tr><td><monologue></td><td>0.25</td><td rowspan="2">0.17</td><td rowspan="3">0.04</td></tr>
<tr><td><statement></td><td>0.11</td></tr>
<tr><td><discussion></td><td colspan="2">0.00</td></tr>
<tr><th>dialogue type</th><th colspan="3">overall summary length</th></tr>
<tr><td><monologue></td><td colspan="3">0.1</td></tr>
</table>

Table 4: The table shows the average summary lengths and the resulting combinations.

If we look at the histogram of the summary lengths on figure 8, we can approximately see a bimodal distribution. This makes sense because they represent the two dialogue types <monologue> and <statement>. The observations between 5% and 20% represent the summary lenghts for the dialogue type <statement> while the observations between 20% and 35% represent the summary lenghts for the dialogue type <monologue>. The red dashed line implies the median summary length for utterances with dialogue type <monologue> and

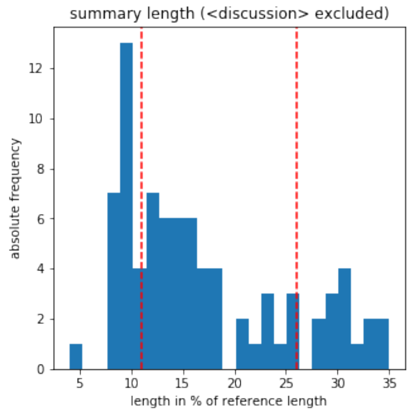

Figure 8: The figure shows the distribution of the summary lengths. There are two groups, which can be assigned to the dialogue type <monologue> (right) and <statement> (left). The red dashed lines represent the mean value of the lengths for the two types.

<statement>. These correspond to 11% and 26% of the original length respectively.

2.3.7 Rouge Score

The evaluation of a summary by a human being is costly and time-consuming. To measure the quality of a summary, there is a performance measure that is widely used. The Recall-Oriented Understudy Gisting Evaluation (Rouge) score has been developed to compare a machine-generated text with a reference summary. The Rouge score is a similarity measure that produces values between 0 and 1, where a value close to 0 implies no similarity and a value close to 1 implies high similarity between the generated summary and the reference summary. In general, the Rouge score gives the ratio between the number of overlapping words and the total number of words in the reference summary. Furthermore, there are different variations of the Rouge score. The Rouge-N score measures different n-gram overlaps such as uni-gram (Rouge-1), bi-gram (Rouge-2) or higher n-grams. The Rouge-L score measures the longest matching sequences of words (Bothe et. al., 2019). However, the Rouge score has limitations in its usefulness, especially for abstract summaries, as the score evaluates summaries based on identical tokens or token combinations. The performance measure is therefore not able to recognise synonyms or evaluate the similarity of the context. Evaluating a summary by a Rouge score therefore only makes limited sense. For this work, the Rouge score is collected without making a statement about the quality of the summary. In any case, the reference summaries are missing for the data used here. For reasons of simplicity, the original text is used as a reference. If we talk about recall in the context of a Rouge score, this indicates how much of the reference summary is covered by the model-generated summary. In other words, recall shows how many of the n-grams of the reference summary also appear in the generated summary. So Recall in the context of Rouge is defined by:

$$ROUGE_{Recall} = \frac{number\ of\ overlapping\ words}{total\ number\ of\ words\ in\ reference\ summary}$$

However, very long summaries can be deceptive in that they cover a lot of common words but do not summarise the content compactly. This is where Precision comes to the help. This measures how much of the model-generated summary is relevant or needed. If the summary contains unnecessary words or is too long, the precision score drops. So Precision in the context of Rouge is defined by:

$$ROUGE_{Precision} = \frac{number\ of\ overlapping\ words}{total\ number\ of\ words\ in\ model\ generated\ summary}$$

The F1-score can then be determined from recall and precision via the harmonic mean (Lin, 2004).

The results in Table 3 show that the precision is close to 1.0 for all dialogue types. This is not surprising as our reference summary is the full text. However, the high precision score indicates that our model creates only few new new n-grams. From the recall score we cannot conclude any information about the quality of the summary. However, it is striking that the values for the Rouge-1 score correspond approximately to the average summary lenghts of the corresponding dialogue types. This can be explained by the fact that the generated summary is limited in its length and thus also the Rouge score, since it is determined here via the original text and not by a reference summary. Furthermore, it is no surprise that the Rouge scores for the dialogue type <monologue> are higher than those for the dialogue type <statement>, as the summaries for the former allow for more detailed texts.

ROUGE-L			
dialogue type	f1-score	precision	recall
<monologue>	0.47	0.99	0.31
<statement>	0.28	0.99	0.16
<monologue>+ <statement>	0.35	0.99	0.22
ROUGE-1			
dialogue type	f1-score	precision	recall
<monologue>	0.39	1	0.25
<statement>	0.2	1	0.11
<monologue>+ <statement>	0.28	1	0.17
ROUGE-2			
dialogue type	f1-score	precision	recall
<monologue>	0.37	0.95	0.23
<statement>	0.18	0.95	0.1
<monologue>+ <statement>	0.26	0.95	0.15

Table 5: The table shows the different ROUGE performance measures for different dialogue types.

3 Conclusion and Outlook

In this work, different approaches have been considered. Not all approaches have led to the desired results. It should also be mentioned that the research question has also evolved during this work. The original idea was to extract action items from a meeting and thereby create a to-do list. This to-do list should then function as a summary. This approach would have been quite promising, as the classification of action items worked very well. However, as the labels of the action items available are too ambiguous and the content of the utterances too domain-specific, it is currently not possible to pursue this approach further. If in the future other data would be available that would make the mentioned limitations obsolete, this approach can be taken up again. However, in the context of dialogues that do not consist of business meetings, such as political debates, it makes little sense to create a to-do list, respectively to classify utterances that contain action items.

With the knowledge gained from the action item classification approach, the idea of supervised text generation was developed. The idea of using a pre-trained model and training it with extractive or abstractive summaries that were available was promising in theory, but here too it became apparent that the available data were too domain-specific and that fine-tuning with a small number of summaries would lead to unsustainable over-fitting of the model. Furthermore, the data comes from simulated scenarios that all contain a given structure, which minimises the variability within the data even more. But if there is enough data including abstractive and/or extractive summaries coming from different domains, it would also be worthwhile to experiment further with this approach.

The approach of assigning utterances to different categories has produced the most promising results in the author's view. The concept presented in this paper leaves it up to the user to decide how detailed the summary should be. A drawback of this approach is that the summaries generated by the T5 model cannot be evaluated via a reference summary. If transcripts are found that have abstractive or extractive summaries, it would be interesting to compare them with the output of this method. However, reading the generated summaries and comparing them with the original utterances shows that from both an objective and subjective point of view, the generated text represents the original statements in an informative and compact way. By extracting the key words, it is also possible for the user to quickly gain an overview of the context of the dialogue. The content of the key words can also be further adapted as required, for example by showing other parts of speech or filtering certain tokens. Furthermore, it would make sense to implement other pre-trained models in the algorithm to compare how they would perform with the model used in this work. One idea that is presented in this approach but could not be fully implemented is topic modelling. It may be helpful for the user to see which topics are discussed in the dialogue and when these topics occur in the dialogue before viewing the transcript summaries. This could be implemented, for example, through an LDA (Tong and Zhang, 2016) applied to individual transcript blocks or neural approaches (Zhang et. al., 2019). Such a segmentation of the dialogue would provide the user an even broader summary of the transcript.

4 References

- J.-M. Torres-Moreno, Automatic Text Summarization, 2014.
- C. Kathri, G. Singh, N. Parikh, Abstractive and Extractive Text Summarization using Document Context Vector and Recurrent Neural Networks, 2018.
- A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. Kaiser, I. Polosukhin, Attention Is All You Need, 2017.
- C. Goo and Y. Chen, Abstractive Dialogue Summarization with Sentence-Gated Modeling Optimized by Dialogue Acts, 2018.
- SpinningBytes AG, https://interscriber.com/, 2021.
- G. Murray and S. Renals, Detecting Action Items in Meetings, 2008.
- M. Purver, P. Ehlen, J. Niekrasz, Detecting Action Items in Multi-Party Meetings: Annotation and Initial Experiments, 2006.
- W. Morgan, P. Chang, S. Gupta, J. M. Brenier, Automatically detecting action items in audio meeting recordings, 2006.
- S. Sheshadri, Identifying Action related Dialogue Acts in Meetings, 2019.
- T. Tran, Francesca Bonin, L. A. Deleris, D. Ganguly, K. Levacher, Preparing a Dataset for Extracting Decision Elements from a Meeting Transcript Corpus, 2018.
- Y. Chen and D. Hakkani-Tür, AIMU: Actionable Items for Meeting Understanding, 2016.
- M. McGregor and J. Tang, More to Meetings: Challenges in Using Speech-Based Technology to Support Meetings, 2017.
- A. Janin, D. Baron, J. Edwards, D. Ellis, D. Gelbart, N. Morgan, B. Peskin, T. Pfau, E. Shriberg, A. Stolcke, C. Wooters, The ICSI meeting corpus, 2003.
- Y. Liu, M. Ott, N. Goyal, J. Du, M. Joshi, D. Chen, O. Levy, M. Lewis, L. Zettlemoyer, V. Stoyanov, RoBERTa: A Robustly Optimized BERT Pretraining Approach, 2019.
- I. Mccowan, G. Lathoud, M. Lincoln, A. Lisowska, W. Post, D. Reidsma, P. Wellner, The AMI Meeting Corpus, 2005.
- G. C. Sergio, https://github.com/gcunhase/AMICorpusXML, 2019.
- M. Lewis, Y. Liu, N. Goyal, M. Ghazvininejad, A. Mohamed, O. Levy, V. Stoyanov, L. Zettlemoyer, BART: Denoising Sequence-to-Sequence Pre-training for Natural Language Generation, Translation, and Comprehension, 2019.
- J. Devlin, M.W.Chang, K. Lee, K. Toutanova, BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding, 2019.

- Y. Wu, M. Schuster, Z. Chen, Q. V. Le, M. Norouzi, W. Macherey, M. Krikun, Y. Cao, Q. Gao, K. Macherey, J. Klingner, A. Shah, M. Johnson, X. Liu, L. Kaiser, S. Gouws, Y. Kato, T. Kudo, H. Kazawa, K. Stevens, G. Kurian, N. Patil, W. Wang, C. Young, J. Smith, J. Riesa, A. Rudnick, O. Vinyals, G. Corrado, M. Hughes, J. Dean, Google's Neural Machine Translation System: Bridging the Gap between Human and Machine Translation, 2016.
- M.BOKAEI, H. SAMETI, Y. LIU, Extractive summarization of multi-party meetings through discourse segmentation, 2015.
- Kaggle, https://www.kaggle.com/headsortails/us-election-2020-presidential-debates/, 2021.
- C. Raffel, N. Shazeer, A. Roberts, K. Lee, S. Narang , M. Matena , Y. Zhou , W. Li, P. J. Liu, Exploring the Limits of Transfer Learning with a Unified Text-to-Text Transformer, 2020.
- C.D. Manning, P. Raghavan and H. Schütze, Introduction to Information Retrieval. Cambridge University Press, pp. 118-120, 2008.
- M. Marcus, B. Santorini, M. Ann Marcinkiewicz, Building a Large Annotated Corpus of English: The Penn Treebank, 1993.
- A. F. Bothe, A. Truesdale, L. Kolbe, State of the Art Summarisation Techniques, 2020.
- C.Y. Lin, ROUGE: A Package for Automatic Evaluation of Summaries, 2004.
- Z. Tong and H. Zhang, A Text Mining Research Based on LDA Topic Modelling, 2016.
- L. Zhang and Q. Zhou, Topic Segmentation for Dialogue Stream, 2019.

5 Appendix

How-To

This section gives instructions on how to run the code for automated dialogue summarisation.

To run the code for unsupervised text generation and key word extraction execute the scripts in the Jupyter Notebook `dialogue_summarizer_avci_2021`.

For technical information read the docstings in the functions of the Jupyter Notebook.

1. Preprocessing data: The first section only serves to clean up the data. Since the data used here is Kaggle data, it is already in a proper format. The dataset should have at least the columns 'speaker', 'minute' and 'text' or similar columns with the given column names.

2. Descriptive Statistics and Visualization: The second section is used to visualise the data, especially the characters, tokens and sentence length in the individual observations. Furthermore, the threshold values are defined in this section, which are decisive for the rest of the code.

3. Text Generation with T5: In this section the pre-trained T5 models are loaded. The given function then summarises the utterances differently according to certain threshold values. Apart from T5 models, other models can also be applied and the hyperparameters of the models, such as the minimum or maximum length of the summaries, can be changed.

4. Key Word Extraction with TF-IDF: This section is dedicated to the allocation of individual discussions into blocks. Then, for the blocks and utterances that have the type monologue or statement, the most important key words are extracted using TF-IDF and merged back with the original data set.

5. Generate EXCEL: In this section, the required information from the dataset such as speakers, summaries and key words are output in an excel sheet. The excel sheet is automatically filtered and colour formatted for the speaker and dialogue types.

6. Evaluation Summary length and Rouge: These sections are devoted to the evaluation of the summaries. First, the lengths of the summaries are compared at different levels. Furthermore, the rouge score for the summaries is determined and their distributions are plotted.

Code and Documentation

Code and documentation can be found on the attached Jupyter Notebooks. Note that only the approach 'Unsupervised Text Generation and Key Word Extraction' has a full documentation.

The Python Version used in this thesis is `Python 3.7.3`.

The models for chapters 2.1 and 2.2 were trained on the Google Colab GPU. All code for chapter 2.3 was computed on a local CPU.

The following Python - Packages have been used in this thesis:

- `warnings`
- `glob`
- `re 2.2.1`
- `datetime`
- `pandas 0.24.2`
- `numpy 1.16.4`
- `matplotlib 3.1.0`
- `seaborn 0.9.0`
- `nltk 3.5`
- `rouge 1.0.0`
- `torch 1.5.0+cpu`
- `transformers 3.4.0`
- `wordcloud 1.8.0`
- `sklearn 0.21.2`
- `gensim 3.4.0`

Summaries for Monologues

speaker	time	key words	summary
Susan Page	00:00	#debate #audience #candidates	it is my honor to moderate this debate, an important part of our democracy. in Kingsbury Hall tonight, we have a small and socially distant audience. everyone in this audience is required to wear a face mask and the candidates will be seated 12 feet apart. the audience is enthusiastic about their candidates, but they've agreed to express that enthusiasm only twice.
Kamala Harris	03:21	#ledger #plan #knew	president's plan is about what we need to do around a national strategy for contact tracing, for testing, for administration of the vaccine, and making sure that it will be free for all. that is the plan that Joe Biden has and that I have, knowing that we have to get a hold of what has been
Mike Pence	04:55	#world #time #decision	from the very first day, president Donald Trump has put the health of America first. he suspended all travel from china, the second largest economy in the world. now, senator Joe Biden opposed that decision.
Mike Pence	05:57	#lives #look #plan	more than 115 million tests had been done to date. we were able to see to the delivery of billions of supplies. under president obama, we believe we'll have literally tens of millions of doses of a vaccine before the end of this year.
Mike Pence	10:11	#people #americans #states	president Trump and I have great confidence in the american people and their ability to take that information into practice. in the height of the epidemic, we surged resources to New Jersey and New York. when the outbreak in the sun Belt happened this summer, again, Americans stepped forward.
Susan Page	13:24	#concerns #vice #president	the president's diagnosis of COVID-19 underscored the importance of the job that you hold. one of you will make history on January 20th. you will be the vice president to the oldest president the united states has ever had.
Mike Pence	14:28	#vaccine #reality #failure	in unheard of time, in less than a year, we have five companies in phase three clinical trials. senator, I just ask you, stop playing politics with people's lives. reality is that we will have a vaccine, we believe, before the end of this year
Mike Pence	15:29	#flu #year #wrong	if the flu had been as lethal as the Coronavirus in 2009, we would have lost 2 million american lives. we still learn from it, and the american people, I'm going to say again, can be proud of what we have done.
Kamala Harris	16:53	#dignity #age #joe	the day I got the call from Joe Biden, it was actually a Zoom call, asking me to serve with him on this ticket was probably one of the most memorable days of my life. we were raised with values that are about hard work, about the value and dignity of public service, and about the importance of fighting for the dignity of all

Kamala Harris	17:58	#woman #department #states	i was elected the first woman of color, and black woman to be elected Attorney General of the state of California. now i serve in the united states Senate as only the second black woman ever elected to the United States Senate. i served on the Senate Intelligence Committee where I've been in regular receipt of classified information about threats
Kamala Harris	21:01	#750 #decisions #debt	we now know Donald Trump owes and is in debt for $400 million. american people have a right to know what is influencing the president's decisions.
Susan Page	23:26	#jobs #growth #taxes	on friday, we learned that the unemployment rate had declined to seven point nine percent in September, but the job growth had stalled. nearly 11 million jobs that existed at the beginning of the year haven't been replaced.
Kamala Harris	25:17	#infrastructure #ll #money	it's about upgrading our roads and bridges, but also investing in clean energy and renewable energy. if you come from a family that makes less than $125,000, you'll go to a public university for free.
Mike Pence	27:29	#jobs #biden #trade	we've already added back 11.6 million jobs because we had a president who cut taxes, rolled back regulation, unleashed american energy and secured four trillion dollars from the Congress of the united states. they want to bury our economy under a two trillion dollar Green New Deal, which you were one of the original co-sponsors of in the United States Senate.
Mike Pence	06:57	#auto #workers #jobs	senator Kamala Harris was one of only 10 members of the Senate to vote against the USMCA. it was a huge win for American auto workers, especially dairy in the upper Midwest.
Mike Pence	08:06	#china #travel #coronavirus	president Trump made the decision before the end of January to suspend all travel from China. he said it was hysterical.
Kamala Harris	10:10	#trade #war #disease	a reputable research firm has done an analysis that shows that leaders of all of our formerly allied countries have now decided that they hold in greater esteem and respect Xi Jinping the head of the Chinese communist party than they do Donald Trump. this is where we are today because of a failure of leadership by this
Mike Pence	14:09	#isis #ve #president	we stood strong with our allies, but we've been demanding. NATO is now contributing more to our common defense than ever before. when president Trump came into office, ISIS had captured an area of the middle east, the size of Pennsylvania.
Kamala Harris	16:49	#strike #feels #hero	what happened to your daughter is awful and it should have never happened, and I know that president Obama feels the same way, but you mentioned soleimani. this is about a pattern of Donald Trump's where he has referred to our men who are serving in our military as suckers and losers.

Mike Pence	21:00	#soleimani #qassem #hesitate	the american people deserve to know Qassem Soleimani, the Iranian general responsible for the death of hundreds of american service members. when the opportunity came, we saw him headed to Baghdad to kill more Americans. the Joe Biden and Kamala Harris actually
Mike Pence	22:10	#faith #knights #hope	our hope is in the hearing next week, unlike Justice Cavenaugh received with treatment from you and others, that we hope she gets a fair hearing. we particularly hope that we don't see the kind of attacks on her Christian faith that we saw before.
Kamala Harris	23:37	#people #issue #election	on the issue of this nomination, Joe Biden and I are both people of faith, and it's insulting to suggest that we would knock anyone for their faith. we are 27 days before the decision about who will be the next president of the united states.
Kamala Harris	24:48	#care #act #coverage	it's the Affordable Care Act like literally in the midst of a public health pandemic. over 210,000 people have died and 7 million people probably have what will be considered a preexisting condition because you contracted the virus. this means that over 20 million people will lose your coverage
Mike Pence	26:41	#court #abortion #judge	Joe Biden and Kamala Harris support taxpayer funding of abortion all the way up to the moment of birth. they want to increase funding to planned parenthood of America.
Kamala Harris	28:26	#lifetime #abraham #lincoln	in 1864, one of the, I think political heroes, certainly the President, is Abraham Lincoln. he was up for reelection and it was 27 days before the election.
Kamala Harris	31:53	#shoulder #arm #life	it brings me to the eight minutes and 46 seconds that America witnessed during which an american man was tortured and killed under the knee of an armed police officer. people around our country marched shoulder to shoulder, arm in arm, fighting for us to finally achieve that ideal of equal justice under
Mike Pence	34:23	#jury #flora #george	justice will be served, but there's also no excuse for the rioting and looting that followed. Flora Westbrook is with us tonight in salt lake city.
Mike Pence	35:27	#enforcement #law #tim	it is remarkable that when senator Tim Scott tried to pass a police reform bill, you filibustered Senator Tim Scott's bill on the Senate floor that would have provided new accountability, new repeat resources. we don't have to choose between supporting law enforcement, proving public safety and supporting our
Kamala Harris	41:13	#work #requirement #bias	we were the first statewide officer to institute a requirement that my agents would wear body cameras and keep them on full-time. we did the work of instituting reforms that were about investing in re-entry. this is the work that we have done and
Kamala Harris	42:31	#democracy #com #integrity	seven members of president George W. Bush's cabinet support our ticket. over 500 generals, retired generals and former national security experts and advisors are supporting our campaign.

Mike Pence	45:44	#clinton #fbi #election	senator, your party has spent the last three and a half years trying to overturn the election results. when Joe Biden was vice president of the united states, the FBI actually spied on president Trump and my campaign. there were documents released this week that the CIA actually made a referral to
Susan Page	47:09	#watch #question #debate	the Utah Debate Commission asked students to write essays about what they would like to ask. I want to close tonight's debate with the question posed by Brecklin Brown.
Mike Pence	48:53	#debate #justice #day	we can debate vigorously as senator Harris and I have tonight. when the debate is over, we come together as Americans. we love a good debate. but we always come together and are always there for one another in times of need.

Table 6: Output for dialogue type <monologue>

Summaries for Statements

speaker	time	key words	summary
Susan Page	01:21	#administration #week #states	39 states have had more COVID cases over the past seven days
Kamala Harris	02:13	#people #workers #affect	210,000 dead people in our country in just the last several months
Susan Page	04:20	#death #population #toll	more than 210,000 americans have died of COVID-19 since
Kamala Harris	07:10	#vice #president #hasn	the vice president knew on January 28th how serious this was.
Mike Pence	08:45	#dr #reality #fauci	the reality is Dr. Fauci said everything that he told the
Mike Pence	11:07	#government #mandates #people	the difference here is President Trump and I trust the american people to make
Kamala Harris	11:58	#people #administration #food	this administration stood on information that, if you had as a
Susan Page	18:58	#information #president #turn	neither president, nor vice president biden has released detailed health information
Mike Pence	19:26	#prayers #concern #forth	the care of the president received at Walter Reed hospital was exceptional.
Mike Pence	20:03	#expressions #congratulate #nomination	the american people have a right to know about the health and well
Mike Pence	22:33	#tens #taxes #president	the president paid tens of millions of dollars in taxes, payroll

Kamala Harris	24:30	#strength #tax #economy	Joe Biden believes you measure the health and the strength of the economy
Mike Pence	26:43	#000 #tax #family	the average american family of four had $2,000 in savings in taxes
Kamala Harris	30:03	#hand #economy #success	the president has reigned over a recession that is being
Kamala Harris	30:28	#bills #barack #hospital	on the one hand, you have Joe Biden who was responsible with
Mike Pence	31:29	#guarantee #healthcare #occasions	Obamacare was a disaster, and the american people remember it well
Mike Pence	32:06	#deal #shape #decline	more taxes, more regulation, banning fracking, abolishing
Susan Page	32:37	#hurricanes #wildfires #change	this year, we've seen record-setting hurricanes in the
Mike Pence	33:15	#conservation #environment #outdoors	our air and land are among the cleanest in the world.
Mike Pence	33:44	#climate #energy #cause	the issue is what's the cause and what do we do about
Mike Pence	34:13	#fracking #ve #countries	the united states has reduced CO2 more than the countries in the Paris
Mike Pence	34:53	#center #management #agreed	president Trump and I believe that forest management has to be front and center
Kamala Harris	36:02	#fact #moody #ban	the american people know that Joe Biden will not ban fracking
Kamala Harris	36:38	#joe #home #crops	the west coast of our country is burning, including my home state of
Kamala Harris	37:07	#science #website #donald	we have seen a pattern with this administration, which is they don
Kamala Harris	37:43	#talk #pride #beings	Joe is about saying we're going to invest in renewable energy
Mike Pence	02:09	#tax #deal #cuts	as I said, Susan, the climate is changing. but once
Mike Pence	02:54	#aoc #resubmit #continue	american people have always cherished our environment and will continue to cherish it
Kamala Harris	03:47	#trade #war #administration	vice-president referred to it as part of what he
Kamala Harris	04:39	#month #end #year	almost half of american renters are worried about whether they're going
Mike Pence	05:18	#china #joe #cheerleader	when Joe Biden was vice-president, we lost 200,000

Susan Page	06:23	#china #relationship #video	we have no more complicated or consequential foreign relationship than the one with china
Kamala Harris	11:59	#relationships #friends #adversaries	we know this in our personal and professional relationships, you got to keep
Kamala Harris	12:41	#intelligence #community #states	the intelligence committee told us Russia interfered in the election of the president
Kamala Harris	13:18	#deal #doesn #donald	the deal has put us in a position where we are less safe
Mike Pence	15:22	#heart #kayla #president	the reality is that when we had an opportunity to save Kayla Mueller
Kamala Harris	18:18	#care #killed #bounty	public reporting that Russia had bounties on the heads of american soldiers
Susan Page	26:00	#care #conditions #president	president says he's going to protect people with pre-exist
Mike Pence	29:45	#powers #separation #court	if you cherish our Supreme Court, you need to reject the Bi
Kamala Harris	30:13	#courts #appointments #court	president Trump appointed 50 people to the court of appeals for lifetime appointments
Susan Page	31:12	#taylor #shot #officers	a 26-year-old emergency room technician was shot and killed
Kamala Harris	33:08	#cops #reform #justice	we need reform of our policing in America and our criminal
Kamala Harris	36:49	#stage #advantage #laws	the only one on this stage has personally prosecuted everything from child
Kamala Harris	37:32	#didn #said #donald	he called Mexicans rapists and criminals;
Mike Pence	38:49	#comments #susan #president	this is one of the things that makes people dislike the media so much
Kamala Harris	43:44	#vote #power #use	we have it within our power in these next 27 days to make the
Mike Pence	44:55	#movement #establishment #americans	president Donald Trump has launched a movement of everyday Americans from every walk
Mike Pence	46:46	#mail #voter #rules	we have a free and fair election. we know we're
Mike Pence	48:13	#news #public #young	we've created literally the freest and most prosperous nation in

Table 7: Output for dialogue type <statement>

Key Words for Discussions

time	discussion block	key words
04:17	<discussion #0>	# thank # president # vice
07:55	<discussion #1>	# mr # seconds # ll
08:02	<discussion #2>	# thank # toilet # kids
08:18	<discussion #3>	# couldn # minute # parents
08:23	<discussion #4>	# people # disservice # months
09:22	<discussion #5>	# thank # vice # president
09:42	<discussion #6>	# people # ability # prayers
09:45	<discussion #7>	# event # administration # row
10:03	<discussion #8>	# people # safety # guidelines
12:33	<discussion #9>	# kamala # help # sorry
12:47	<discussion #10>	# americans # half # anthony
13:11	<discussion #11>	# professionals # absolutely # line
16:32	<discussion #12>	# month # disability # safeguards
20:30	<discussion #13>	# stakes # stage # reality
20:46	<discussion #14>	# challenge # candidates # voters
22:01	<discussion #15>	# susan # joe # honest
22:22	<discussion #16>	# thank # vice # respect
23:16	<discussion #17>	# segue # energy # kamala
26:09	<discussion #18>	# welcome # thinks # benefit
26:26	<discussion #19>	# president # report # comeback
28:51	<discussion #20>	# history # year # pence
28:56	<discussion #21>	# speaking # truth # fact
29:14	<discussion #22>	# tax # cuts # break
29:35	<discussion #23>	# mind # don # conversation
29:39	<discussion #24>	# joe # recession # fracking
30:54	<discussion #25>	# protections # conditions # just
31:06	<discussion #26>	# crosstalk # disease # cancer
35:24	<discussion #27>	# president # hurricanes # goods
35:36	<discussion #28>	# deal # framework # campaign
38:16	<discussion #29>	# threat # change # president
03:40	<discussion #30>	# vice # environment # care
05:50	<discussion #31>	# president # vice # growth
06:02	<discussion #32>	# jobs # thank # mandates
09:03	<discussion #33>	# vice # president # field
09:13	<discussion #34>	# competitors # adversaries # relationship
09:29	<discussion #35>	# president # office # loss
11:33	<discussion #36>	# ve # definition # role
14:03	<discussion #37>	# thank # vice # president
16:20	<discussion #38>	# president # donald # trump
16:40	<discussion #39>	# thank # topics # course

18:53	<discussion #40>	# issues # security # democracy
19:16	<discussion #41>	# ve # president # slanders
19:31	<discussion #42>	# states # deployed # daughters
19:54	<discussion #43>	# rules # race # enforce
20:14	<discussion #44>	# topic # court # senate
20:32	<discussion #45>	# abortion # wade # roe
21:56	<discussion #46>	# judge # place # supreme
23:17	<discussion #47>	# california # senator # restrictions
25:47	<discussion #48>	# coverage # let # medicare
27:42	<discussion #49>	# seats # rules # question
28:03	<discussion #50>	# court # question # biden
28:18	<discussion #51>	# packing # lesson # way
29:40	<discussion #52>	# talk # let # yeah
31:04	<discussion #53>	# time # senator # thank
34:08	<discussion #54>	# vice # taylor # breonna
36:38	<discussion #55>	# record # investments # unemployment
37:16	<discussion #56>	# supremacists # separates # diversity
38:46	<discussion #57>	# minute # glad # record
39:42	<discussion #58>	# francisco # blacks # incarceration
40:04	<discussion #59>	# justice # thank # sir
40:24	<discussion #60>	# issue # opportunity # seconds
40:39	<discussion #61>	# holes # model # appreciate
41:06	<discussion #62>	# thank # commitment # points
42:03	<discussion #63>	# transfer # election # president
44:31	<discussion #64>	# power # transfer # president
49:51	<discussion #65>	# leaders # words # perspective
50:17	<discussion #66>	# joe # run # division
50:44	<discussion #67>	# joe # dignity # suffering
51:10	<discussion #68>	# leadership # vote # person
51:37	<discussion #69>	# thank # senator # vice

Table 8: Output for diaogue type <discussion>