

19.52

Practical applications
of Artificial Intelligence
in business

9-10th June / Conference Online
11th June / Workshops Online



Text Summarization with Transformer Models – Exploring New Frontiers in NLP

Natalia Kees
qdive GmbH

www.aisummit.today

My Background



- Originally from Ukraine, live and work in Germany
- Political and Computer Scientist
- Focus on NLProc
- Data Scientist at qdive GmbH, core services:
 - **Guide:** Identify **potentials** for data products
 - **Build:** Use prototypes for **exploration** of solutions and bring these as data products into **production**
 - **Platform:** Data **platform** with Site Reliability Engineering (DevOps)



Today's Talk

1. Text Summarization
 - Motivation
 - Quality Criteria
2. Methods
 - Introduction to Bidirectional Transformer Models
 - Model
3. Summarization of Long Texts
 - Solution 1: Long Range Sequence Modeling
 - Solution 2: Heuristics
4. Evaluation
 - The Data
 - Results
5. Summary

Text Summarization

Motivation

I entered into a payment arrangement with the debt collector, with my first payment due approximately 23 September 2018, in order to attempt to bring my account balance up to date and avoid further garnishment against my account. I had attempted to enter into payment arrangements previously to avoid garnishment, however the company told me every time I attempted that they would not allow me to make any payments on my account without providing all of the personal information they requested from me and I without having a job that I receive a paycheck from. I worked independently as a A NAMED ENTITY so I did not have payroll, I could only bring cash, so I wasnt allowed to make payments on my accounts at several points throughout the last several years. I entered into the payment arrangement and my first payment was September 2018. I was never given any copies of my agreement, as I was told that they do not provide this to people except over the phone, but the staff over the phone on a recorded call that they would no longer send any garnishments against me as long as I made a good faith effort to pay. I was told the only day I could set my payment arrangement up to be due on (on a weekly basis). I explained that I was concerned about payments being due on Mondays, unable to make it to the office sometimes on Mondays. The staff member told me that there was nothing else possible, and that my only choice was to be Mondays, but that there is a three-day grace period available just in case I couldnt make it on that day. Additionally, I have explained to them several times funds available in my bank account and the times they levied my bank account were from child support, that I could prove, and that the funds they kept in account were exempt funds. These funds were never returned. I began been making payments regularly, and there was a week that my payment was due NAMED ENTITY) and I didnt make it into their office on that week. My mother had been admitted to the hospital, and between the holiday and her hospital unavailable to make it to their office. I made it as soon as I could to make my next payment and explained what happened to the staff member, and I experienced was late in getting my payment to them and wasnt sure if that messed up my payment arrangement or not. Because I was never sent any information about payment arrangement, I was not able to review this information to figure out what to do at the time and I didnt have phone service available at the hospital this information to the staff member in the office, and I was told that there shouldnt be any issues with my payment arrangement and to continue sending scheduled, which I did. On 23 September 2018 I realized that the funds in my account were gone after my card was declined at a merchant and I called to notified me that a bank levy had been sent, and was received by the bank on 23 September 2018, which was the day after my last payment made to this actively making a good faith effort to pay my account, and the company sent a levy to my account anyways and sent my account into the negative by \$1 draining the \$32.00 that was in it. Those funds were child support funds, which I reported to the company today. I noticed today (23 September 2018) the negative \$110.00, which made it impossible for me to receive another payment for child support in my account without the funds being taken up by the bank. When I called the company to ask why they garnished my account after previously promising not to and telling me that my payment arrangement wasnt late when my mother was hospitalized (I am her legal Power of Attorney), I was told that there was no payment arrangement on file, that they had noted they had previously informed me that I needed to call and set up a new one, after they had told me a different time that I wouldnt be allowed to set up any arrangements, which is false. She also told me that there was no grace period allowed, and when I confronted her about what the other staff member had said three day grace period, she then recanted her story and stated that there IS in fact a grace period available, but then stated it is more for when a payment weekend. She then told me if I wanted to take up any concerns about child support funds being garnished which should be exempt, that I had to do so in this account, and that my bank are the only ones able to tell them that the funds in my account are exempt funds. This is not true. This phone call between me and the company are recorded, as well as the phone call I made with the bank informing me that they have nothing to do with informing the creditor that the funds are exempt. To date I have been given no notice to dispute the garnishment or inform them that the funds are exempt, and this is a frequent occurrence with experience and the experiences of thousands of other consumers in our town, the creditor, who is owned and operated under the name of an attorney, does people to even make reasonable attempts to pay their account balances without meeting extensive requirements, will deny attempts to pay accounts, using misleading statements regularly and illegally misrepresents the law and what people need to do in order to take care of their debts, frequently garnishes known to be exempt forcing consumers to go through extensive legal processes and fees in order to dispute the funds being garnished and claim them as exempt, will repeatedly send garnishments to accounts that they know have little to no money in them week after week, sending consumers into more and more debt with their banks as they repeatedly incur huge legal order processing fees and can not even maintain an open bank account long enough to save any money to pay the debts off in the first place, does not provide accurate documentation of payment arrangements or account activity, lies about payment arrangements, promises to not garnish accounts and then garnishes them anyways, and frequently communicates court summons to debt collection court for status hearings by post card, which is also illegal. This creditor has done this for years and should be penalized for their procedural harassment and bullying of creditors and denying people the ability to pay their debts. By law a consumer doesnt have to provide any more information than what they are comfortable with sharing, and the company uses this to deny people the ability to pay their debts. They do not cease collection efforts when notified that funds are exempt or that debts are disputed, and they frequently fail to give notice prior to legal proceedings which ensure an individual does not have time to attend the hearing to dispute the debt in the first place, resulting in default judgement. On behalf of almost everyone in A NAMED ENTITY , KS, please hold this creditor accountable for their malicious, negligent, and bullying debt collection practices. This is not right for anyone to have to be subjected to this kind of nonsense and I have the calls recorded finally to try to help prove it. We all believe that this company is actually embezzeling through the abuse of state statutes and abusive collection practices, and that they are somehow involved with the state in the funneling of funds gained by nickel and diming people in small enough amounts that they dont have the opportunity to contest it without incurring huge financial or time losses to try to do so.

I entered into a payment arrangement with the debt collector. The company told me every time that they would not allow me to make any payments on my account without providing all of the personal information they requested from me and I without having a job that I receive a paycheck from. I did not have payroll. I noticed today (23 September 2018) that my account was negative \$110.00, I was told information which is false. This creditor has done this for years and should be penalized for their procedural harassment and bullying of creditors and denying people the ability to pay their debts.

Text Summarization

Motivation

I entered into a payment arrangement with the debt collector, with my first payment due approximately 23 September 2018, in order to attempt to bring my account balance up to date and avoid further garnishment against my account. I had attempted to enter into payment arrangements previously to avoid garnishment, however the company told me every time I attempted that they would not allow me to make any payments on my account without providing all of the personal information they requested from me and I without having a job that I receive a paycheck from. I worked independently as a NAMED ENTITY so I did not have payroll, I could only bring cash, so I wasnt allowed to make payments on my accounts at several points throughout the last several years. I entered into the payment arrangement and my first payment wa September 2018. I was never given any copies of my agreement, as I was told that they do not provide this to people except over the phone, but the staff over the phone on a recorded call that they would no longer send any garnishments against me as long as I made a good faith effort to pay. I was told the only day I could set my payment arrangement up to be due on (on a weekly basis). I explained that I was concerned about payments being due on Monday unable to make it to the office sometimes on Mondays. The staff member told me that there was nothing else possible, and that my only choice was to be Mondays, but that there is a three-day grace period available just in case I couldnt make it on that day. Additionally, I have explained to them several times funds available in my bank account and the times they levied my bank account were from child support, that I could prove, and that the funds they kept I account were exempt funds. These funds were never returned. I began been making payments regularly, and there was a week that my payment was due NAMED ENTITY) and I didnt make it into their office on that week. My mother had been admitted to the hospital, and between the holiday and her hospital unavailable to make it to their office. I made it as soon as I could to make my next payment and explained what happened to the staff member, and I expi was late in getting my payment to them and wasnt sure if that messed up my payment arrangement or not. Because I was never sent any information ab payment arrangement, I was not able to review this information to figure out what to do at the time and I didnt have phone service available at the hospital this information to the staff member in the office, and I was told that there shouldnt be any issues with my payment arrangement and to continue sending scheduled, which I did. On 23 September 2018 I realized that the funds in my account were gone after my card was declined at a merchant and I called to notified me that a bank levy had been sent, and was received by the bank on 23 September 2018, which was the day after my last payment made to this actively making a good faith effort to pay my account, and the company sent a levy to my account anyways and sent my account into the negative by \$1 draining the \$32.00 that was in it. Those funds were child support funds, which I reported to the company today. I noticed today (23 September 2018) the negative \$110.00, which made it impossible for me to receive another payment for child support in my account without the funds being taken up by the n When I called the company to ask why they garnished my account after previously promising not to and telling me that my payment arrangement wasnt late when my mother was hospitalized (I am her legal Power of Attorney), I was told that there was no payment arrangement on file, that they had noted they had previously informed me that I needed to call and set up a new one, after they had told me a different time that I wouldnt be allowed to set up any arrangements, which is false. She also told me that there was no grace period allowed, and when I confronted her about what the other staff member ha three day grace period, she then recanted her story and stated that there IS in fact a grace period available, but then stated it is more for when a payme weekend. She then told me if I wanted to take up any concerns about child support funds being garnished which should be exempt, that I had to do so th account, and that my bank are the only ones able to tell them that the funds in my account are exempt funds. This is not true. This phone call between m company are recorded, as well as the phone call I made with the bank informing me that they have nothing to do with informing the creditor that the func exempt. To date I have been given no notice to dispute the garnishment or inform them that the funds are exempt, and this is a frequent occurrence with experience and the experiences of thousands of other consumers in our town, the creditor, who is owned and operated under the name of an attorney, di people to even make reasonable attempts to pay their account balances without meeting extensive requirements, will deny attempts to pay accounts, u misleading statements regularly and illegally misrepresents the law and what people need to do in order to take care of their debts, frequently garnishes know to be exempt forcing consumers to go through extensive legal processes and fees in order to dispute the funds being garnished and claim them as exempt, will repeatedly send garnishments to accounts that they know have little to no money in them week after week, sending consumers into more and more debt with their banks as they repeatedly incur huge legal order processing fees and can not even maintain an open bank account long enough to save any money to pay the debts off in the first place, does not provide accurate documentation of payment arrangements or account activity, lies about payment arrangements, promises to not garnish accounts and then garnishes them anyways, and frequently communicates court summons to debt collection court for status hearings by post card, which is also illegal. This creditor has done this for years and should be penalized for their procedural harassment and bullying of creditors and denying people the ability to pay their debts. By law a consumer doesnt have to provide any more information than what they are comfortable with sharing, and the company uses this to deny people the ability to pay their debts. They do not cease collection efforts when notified that funds are exempt or that debts are disputed, and they frequently fail to give notice prior to legal proceedings which ensure an individual does not have time to attend the hearing to dispute the debt in the first place, resulting in default judgement. On behalf of almost everyone in A NAMED ENTITY , KS, please hold this creditor accountable for their malicious, negligent, and bullying debt collection practices. This is not right for anyone to have to be subjected to this kind of nonsense and I have the calls recorded finally to try to help prove it. We all believe that this company is actually embezzling through the abuse of state statutes and abusive collection practices, and that they are somehow involved with the state in the funneling of funds gained by nickel and diming people in small enough amounts that they dont have the opportunity to contest it without incurring huge financial or time losses to try to do so.

I entered into a payment arrangement with the debt collector. The company told me every time that they would not allow me to make any payments on my account without providing all of the personal information they requested from me and I without having a job that I receive a paycheck from. I did not have payroll. I noticed today (23 September 2018) that my account was negative \$110.00, I was told information which is false. This creditor has done this for years and should be penalized for their procedural harassment and bullying of creditors and denying people the ability to pay their debts.

- Customer communication processing
 - Assistance for other machine learning problems
 - Data augmentation (e.g. Question Answering)

Quality Criteria

Factual Correctness

Relevance

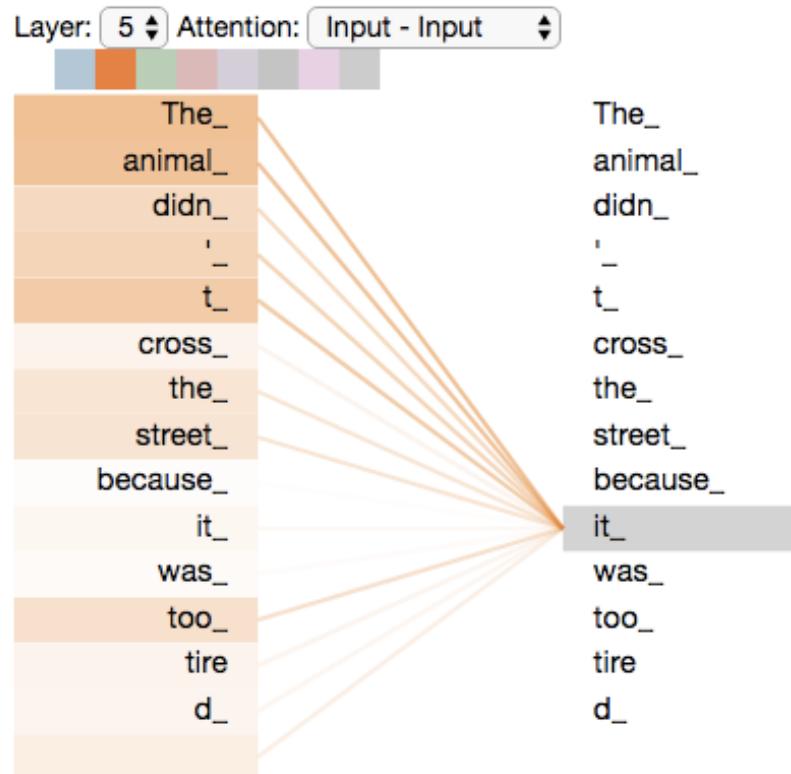
Good Readability

Intelligibility

Adequate Level of Abstractiveness

I entered into a payment arrangement with the debt collector. The company told me every time that they would not allow me to make any payments on my account without providing all of the personal information they requested from me and I without having a job that I receive a paycheck from. I did not have payroll. I noticed today (23 September 2018) that my account was negative \$110.00, I was told information which is false. This creditor has done this for years and should be penalized for their procedural harassment and bullying of creditors and denying people the ability to pay their debts.

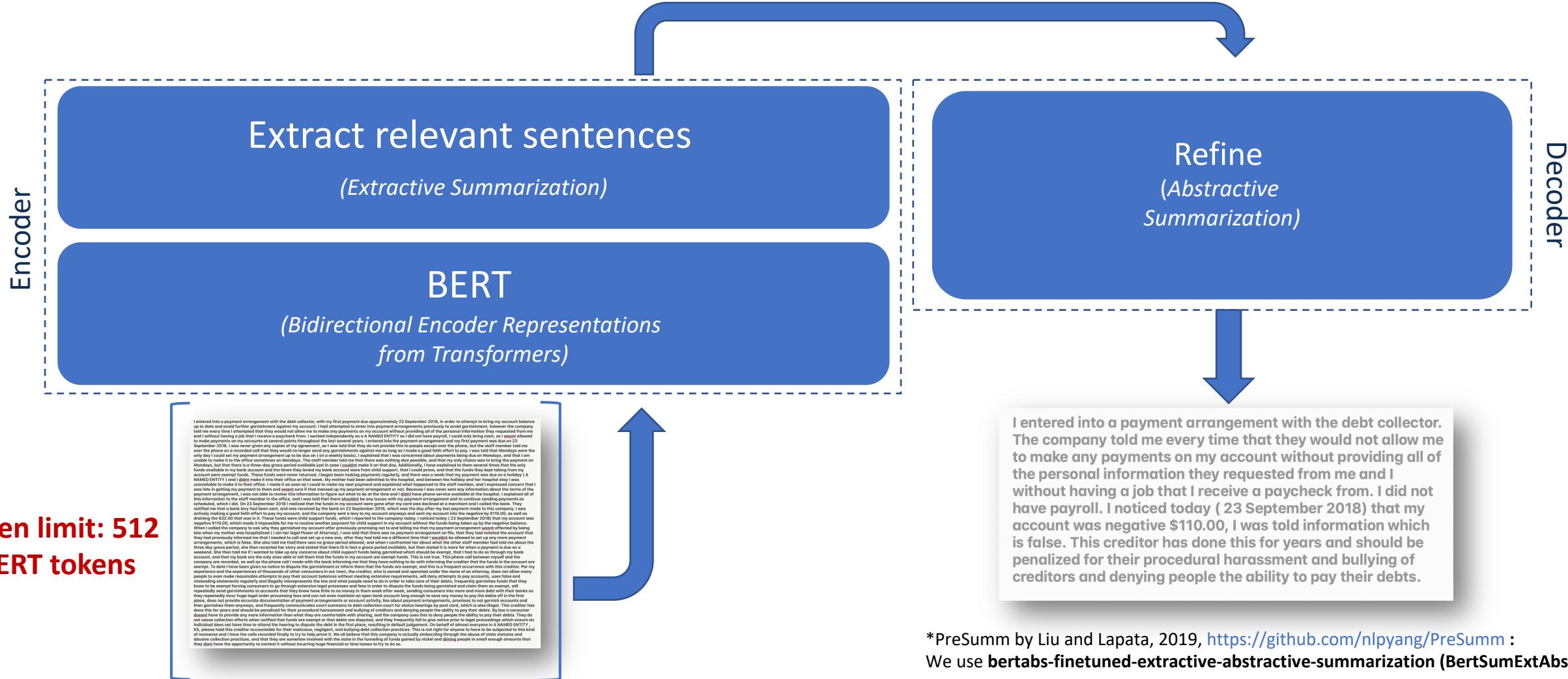
Introduction to Bidirectional Transformer Models



- Transformer architectures with multi-headed **(self-)attention**
- Capture **context-specific** facets of meanings of words and **semantic relations** between them thanks to self-attention
- **Speed up** training dramatically
- Easy to **fine-tune** on downstream tasks
- Require much **less training data** for each specific case
- **Outperform** competitors by a large margin
- Enable high levels of **abstractiveness**
- Examples: **BERT, RoBERTa, DistilBERT, ...**

Source: <http://jalammar.github.io/illustrated-transformer/>

Example Model (BertSumExtAbs*)



Token limit: 512 BERT tokens

I entered into a payment arrangement with the debt collector. The company told me every time that they would not allow me to make any payments on my account without providing all of the personal information they requested from me and I without having a job that I receive a paycheck from. I did not have payroll. I noticed today (23 September 2018) that my account was negative \$110.00, I was told information which is false. This creditor has done this for years and should be penalized for their procedural harassment and bullying of creditors and denying people the ability to pay their debts.

*PreSumm by Liu and Lapata, 2019, <https://github.com/nlpyang/PreSumm> :
We use **bertabs-finetuned-extractive-abstractive-summarization** (**BertSumExtAbs**)

Solution 1: Long Range Sequence Modeling

- Main goal: reducing the complexity of the self-attention from quadratic to linear
- Possible ways to reach this:
 - Local attention (e.g. Longformer)
 - Memory-compressed attention through recurrence (e.g. Compressive Transformer)
 - High-dimensional data compression (e.g. Linformer)

Source: <https://huggingface.co/blog/long-range-transformers>

Solution 1: Long Range Sequence Modeling

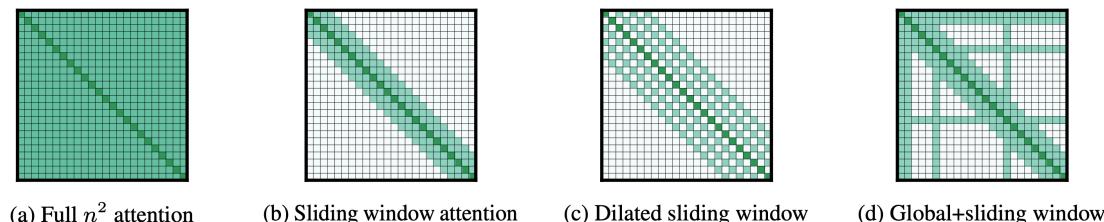
Example: Local Attention with Longformer

■ How it works

- Replacing conventional self-attention with a combination of windowed/local/sparse attention and global attention
- Implementation by simply replacing the standard self-attention in the code
- Can be used on top of pre-trained checkpoints with standard attention for fine-tuning
- Sequence length: 4096

■ For summarization

- Longformer-Encoder-Decoder (LED):
 - Encoder: Sliding window local attention + Global attention
 - Decoder: full attention



(a) Full n^2 attention

(b) Sliding window attention

(c) Dilated sliding window

(d) Global+sliding window

Source: Beltagy et al. 2020, <https://arxiv.org/pdf/2004.05150.pdf>

Solution 1: Long Range Sequence Modeling

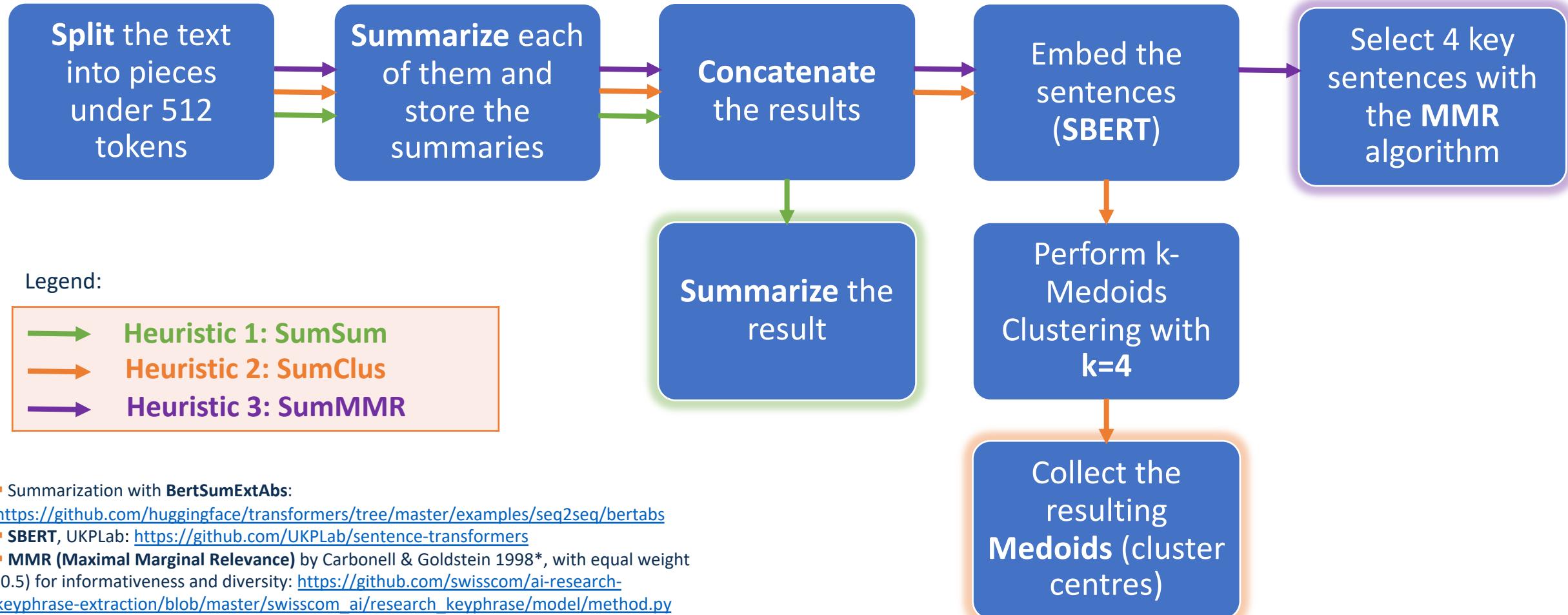
- Takes time to adapt, no out-of-the-box usage
- GPU-intensive

*Source: Tay et al. 2020, <https://arxiv.org/pdf/2011.04006.pdf>

Solution 2: Heuristics

- Use an out-of-the-box **fully trained model**, e.g. PreSumm, trained initially on short texts on generic data
- Use heuristics based on **splitting long texts and produce good summaries** irrespectively of the training procedure
- Benefits:
 - Are quick and easy to perform, example implementation: <https://github.com/nkees/long-text-summarization>
 - Can work as a baseline for comparing your adapted methods against

Solution 2: Heuristics



The Data

An official website of the United States government

Español 中文 Tiếng Việt 한국어 Tagalog Русский العربية Kreyòl Ayisyen (855) 411-2372

cfpb Consumer Financial Protection Bureau

Consumer Tools ▾ Practitioner Resources ▾ Data & Research ▾ Policy & Compliance ▾ About Us ▾

Search Submit a Complaint

Consumer Complaint Database

This database is a collection of complaints about consumer financial products and services that we sent to companies for response.



Things to know before you use the database

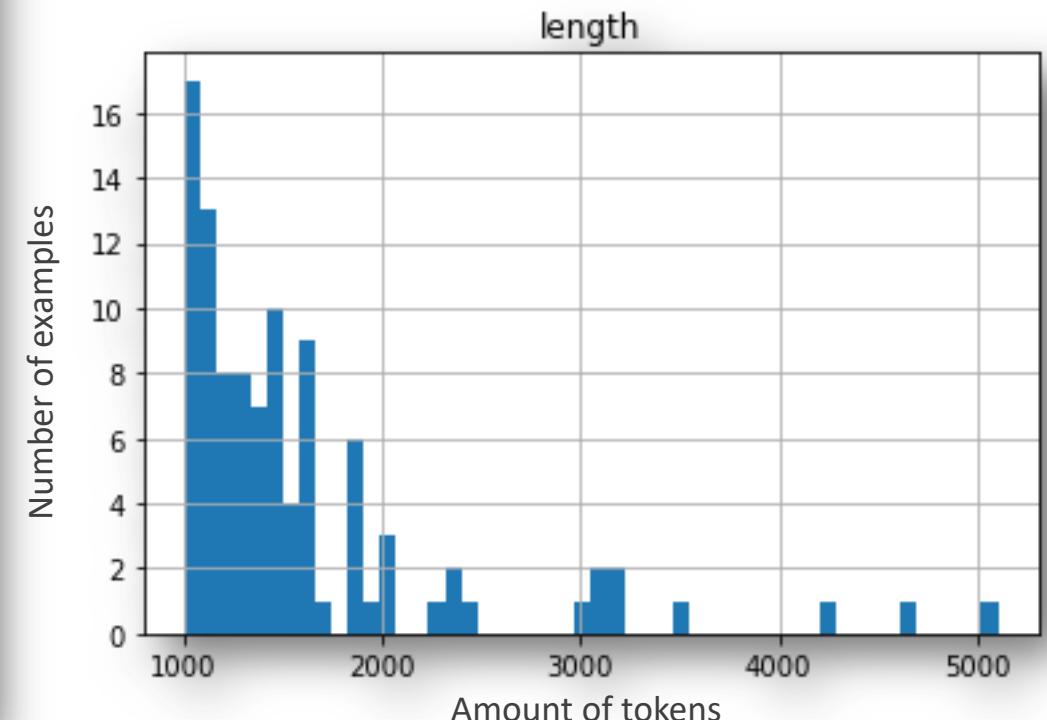
Complaints are published after the company responds, confirming a commercial relationship with the consumer, or after 15 days, whichever comes first. [Learn more](#)

Complaints are not necessarily representative of all consumers' experiences with a financial product or company. [Learn more](#)

We don't verify all the allegations in complaint narratives. [Learn more](#)

Source: <https://www.consumerfinance.gov/data-research/consumer-complaints/>

Distribution of narratives by length



Results

	Heuristic 1: SumSum				Heuristic 2: SumClus				Heuristic 3: SumMMR			
	rouge-1	rouge-2	rouge-3	rouge-L	rouge-1	rouge-2	rouge-3	rouge-L	rouge-1	rouge-2	rouge-3	rouge-L
F1	0.368	0.149	0.104	0.300	0.414	0.185	0.133	0.308	0.411	0.176	0.127	0.295
Precis.	0.436	0.177	0.126	0.344	0.374	0.166	0.119	0.279	0.369	0.158	0.114	0.265
Recall	0.347	0.140	0.098	0.285	0.506	0.233	0.172	0.369	0.500	0.214	0.156	0.352
Len. Rat.	0.897				1.838				1.733			

- **SumClus** (heuristic 2) shows best results in terms of **quantitative comparison of F1-Scores**
- **SumSum** (heuristic 1) performs best in terms of **precision and length ratio, approximating human-made summaries convincingly**

Summary

- Powerful encoders, which bidirectional encoders are, allow to reach human-like summaries with good **readability** and high **relevance** of contents
- Working with **long texts** is a complicated, but **feasible**, task
- There are ways to change the model's architecture to encode long texts efficiently, like with the use of **Longformers**
- One can also make use of some **heuristics** to provide good summaries for long texts **without** the need to adjust the model architecture



<https://github.com/nkees/long-text-summarization>

qdive
passion for data

Thank you for your attention!

Natalia Kees

Data Scientist

qdive GmbH

Leopoldstr. 28a

80802 Munich

Germany

Office 0049 89 455990

Mobile 0049 160 91334718

eMail natalia.kees@qdive.io

Internet www.qdive.io