

Text Classification and Summarization

Fortnightly Innovation Session

Sue Liu, 20th September 2019

Consumer complaints revisited

- Automatic text summarization
- Applications: back to complaint classification
- What's next?

More complaints

- US consumer complaints dataset (<https://catalog.data.gov/dataset/consumer-complaint-database>)
- Updated daily - data now up to 2nd September: ~430,000 complaints with narratives
- Label consolidation

Credit reporting → **Credit reporting, credit repair services, or other personal consumer reports**

Credit card

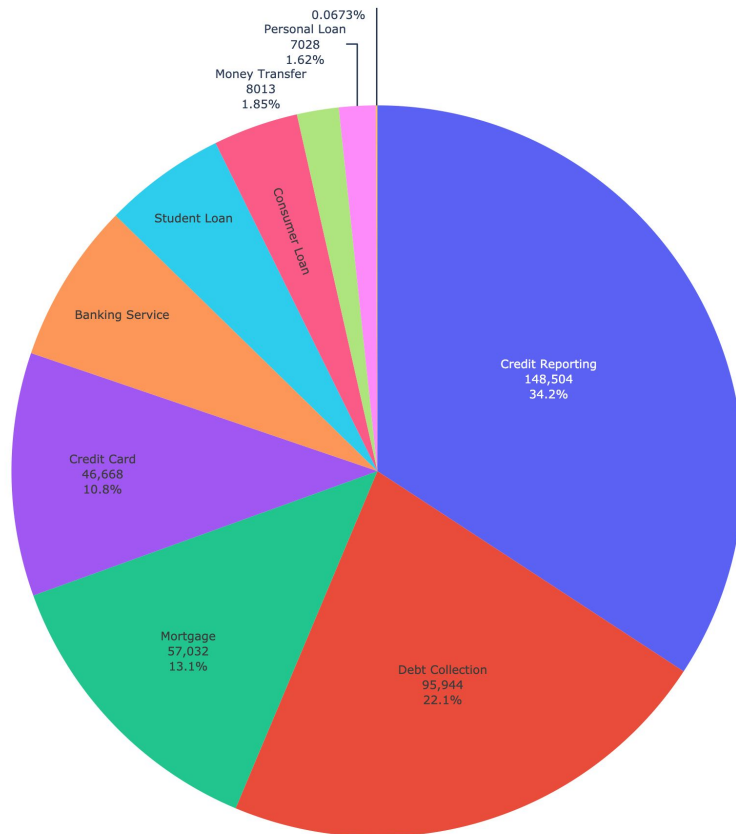
Prepaid card → **Credit card or prepaid card**

Money transfers

Virtual currency → **Money transfer, virtual currency, or money service**

Vehicle loan or lease → **Consumer Loan**

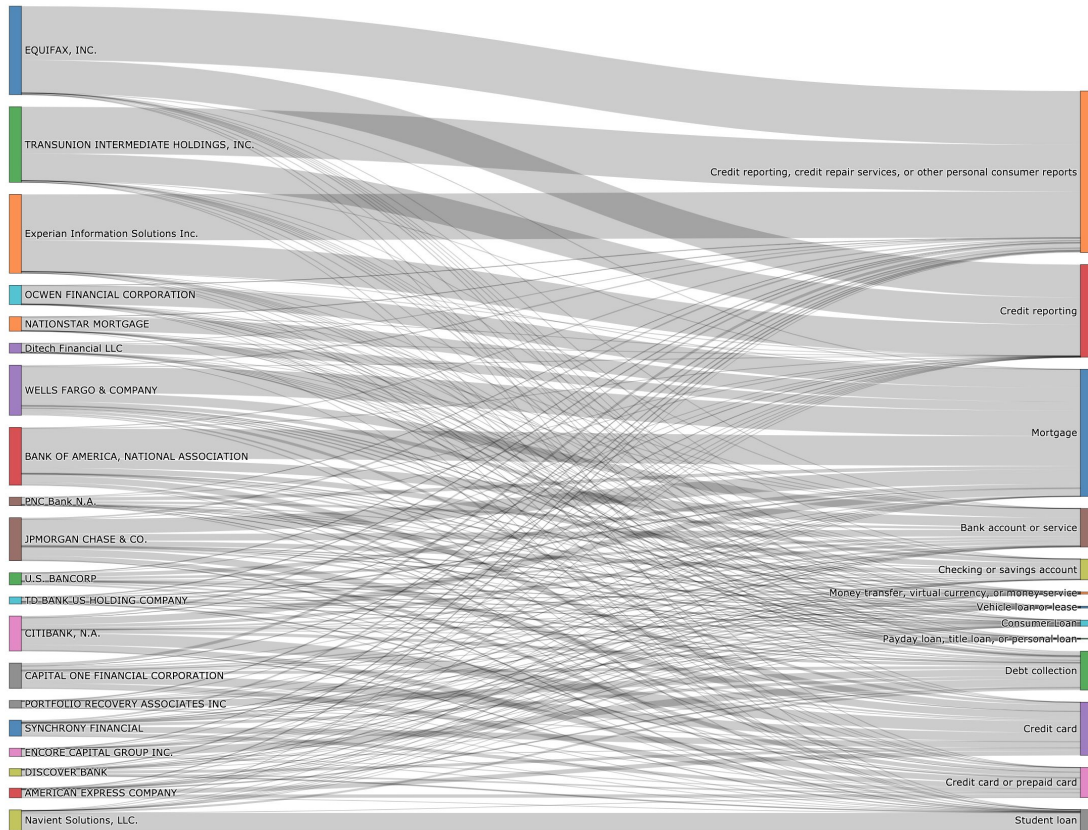
Payday loan → **Payday loan, title loan, or personal loan**



Complaint type by company

Companies

Complaint type



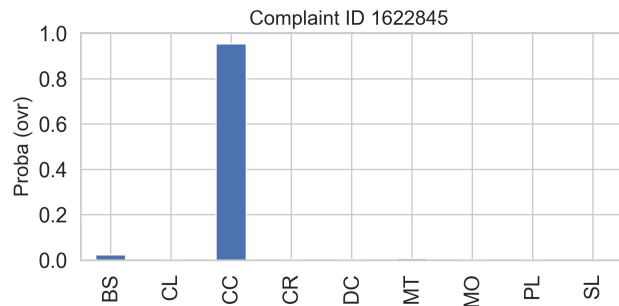
Product Classification

- Re-run of Julia's work with new data / labels
- Predict product category purely based on complaint narratives
- Train on 277596 , validate on 69400 complaints
- Bag-of-words model with Tf-Idf vectorization Scikit-learn classifiers
 - Logistic Regression with 5-fold CV
 - LinearSVC
- Simple LSTM
 - Embedding dim = 100
 - Max number of words = 250
 - Top 50000 words
- Loans / Debts are difficult to predict (one often implies the other!)

Scaled recall matrix

Model	Accuracy									
Logistic Regression	0.79									
Linear SVC	0.81									
LSTM	0.82									
Banking Service	0.78	0.02	0.04	0.00	0.00	0.31	0.01	0.08	0.00	
Consumer Loan	0.01	0.68	0.01	0.01	0.01	0.02	0.01	0.30	0.01	
Credit Card	0.09	0.05	0.79	0.01	0.02	0.18	0.00	0.11	0.01	
Credit Reporting	0.05	0.39	0.15	0.76	0.13	0.05	0.06	0.28	0.14	
Debt Collection	0.03	0.16	0.06	0.05	0.77	0.08	0.03	0.60	0.09	
Money Transfer	0.03	0.00	0.01	0.00	0.00	0.82	0.00	0.01	0.00	
Mortgage	0.02	0.05	0.01	0.01	0.01	0.05	0.90	0.18	0.02	
Payday Loan	0.01	0.05	0.01	0.00	0.00	0.01	0.00	0.68	0.01	
Student Loan	0.00	0.02	0.00	0.00	0.01	0.01	0.00	0.10	0.89	
	Banking Service	Consumer Loan	Credit Card	Credit Reporting	Debt Collection	Money Transfer	Mortgage	Payday Loan	Student Loan	

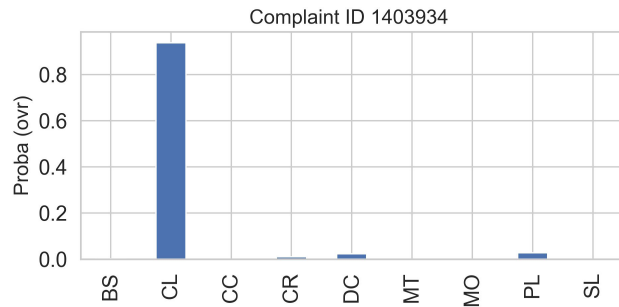
Misclassifications



Actual: Bank account or service

Predicted: Credit card or prepaid card (0.96)

Narrative: I am one of the many affected by the RushCard fiasco. Today is the XX/XX/2015. My card is still being declined. I spent XXXX hrs on hold today. I have received txt saying my funds are on my old expired card.



Actual: Debt collection

Predicted: Consumer loan (0.94)

Narrative: My son and I were on an auto loan together and he couldn't afford it anymore so we voluntarily surrendered the vehicle to the bank. we were told they would send a final bill after car is sold in an auction and haven't heard anything yet from them since we gave car back in XXXX 2015. Yesterday XXXX 2015 they somehow got my tenants number illegally and called her wanting her to disclose our personal information to them, like where we work and our phone numbers etc. And they also told her that we both may get a warrant for our arrest ... they have the car back so how can they threaten to arrest us for something we no longer poses? This bank has major issues and needs to be stopped..they had no right contacting my tenant and involving them and we don't know how they got her number except googling it which should be illegal to do in this scenario. They also falsely reported on our credit reports that the car was a repossession not a voluntary surrender like it was.

Verbose complainers

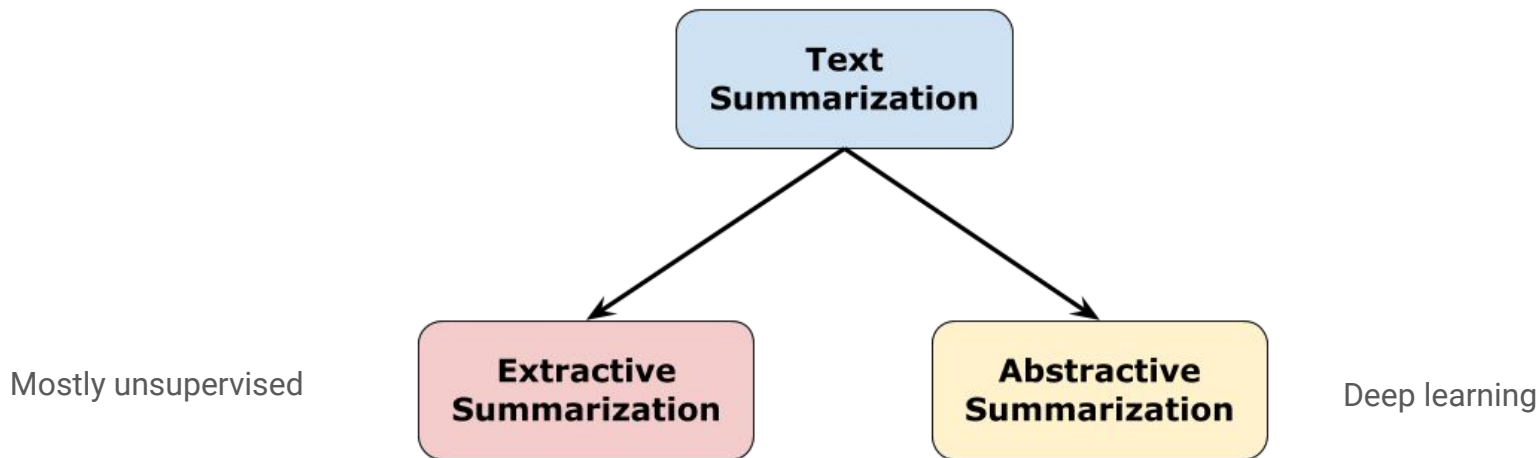
'Hello, we have big problem again, there is a foreclosure sale scheduled for our home on XX/XX/XXXX and we are out of options and we have exhausted ourselves with this horrible company. We have had many attempts to receive an affordable monthly payment but no resolution and no affordable monthly payment offer has been given. We have sent in a request for mortgage assistance form along with bank statements taxes paystubs and other documents necessary for a loan modification review. We recently sent Mr. Cooper all documents but Mr. Cooper ahs the sale still scheduled. \nAs part of this bailout, on XX/XX/XXXX, the United States Government took over XXXX XXXX and pumped billions, which in XX/XX/XXXX, then Treasury Secretary XXXX XXXX said as much as {XXXX} XXXX would be pumped into XXXX XXXX and XXXX XXXX, each, which was separate from the {\$700.00} in TARP funds given out to other lenders. \n\n1. The guidelines issued by the United States Treasury set forth a detailed process whereby a participating servicer, either action or not acting through its subsidiary, must : a. identify loans that are subject to modification under the HAMP program, both through its own review and in response to requests for modification from individual homeowners ; b. collect financial and other personal information from the homeowners to evaluate whether the homeowner is eligible for a loan modification under HAMP ; c. institute a modified loan with a reduced payment amount as per a mandated formula, that is effective for a three-month trial period for borrowers that are eligible for a modification ; and d. provide a permanently modified loan to those homeowners who comply with the requirements during the trial period. Whether the homeowner qualifies for a modification or not, participating servicers are also required to provide written notices to every mortgage borrower that has been evaluated for a loan modification, whether or not the borrower has been found eligible. \n\nHAMP and its associated directives also set prohibitions against certain conduct including demanding upfront payments in order to be evaluated for a loan modification, instituting or continuing foreclosures while a borrower is being evaluated for a loan modification, assigning a single point of contact, maintaining an adequate level communication with borrowers, maintaining proper caseload to ensure HAMP objections are met, and restrictions on the way a servicer may report the borrower to credit reporting agencies. \n\nAll servicers have systematically failed to comply with the terms of the HAMP directives and has regularly and repeatedly violated several of its prohibitions. Now that HAMP is over and expired due to servicers prolonging these modifications they are now proceeding with other alternatives. \n\nUnder HAMP, the federal government incentivizes participating servicers to make adjustments to existing mortgage obligations in order to make the monthly payments more affordable. Servicers receive at least {\$1000.00} for each HAMP modification. However, this incentive is countered by a number of financial factors that make it more profitable for a mortgage servicer avoid modification and to continue to keep a mortgage in a state of default or distress and to push loans toward foreclosure. This is especially true in cases where the mortgage is owned by a third-party investor and is merely serviced by Service companies because Service companies do not carry a significant risk of loss in the event of foreclosure. \n\nCalifornia Business and Professions Code 17200 et seq., also known as the California Unfair Competition Law (UCL), prohibits acts of unfair competition, including any unlawful, unfair, or deceptive business act or practice as well as unfair, deceptive, untrue or misleading advertising. Mr. Cooper and all other previous servicers and their Investors conduct was unlawful in that : They illegally began and prosecuted the non-judicial foreclosure based on the lien to the 1st Deed of Trust, despite i was not in breach of payments on said lien ; They engaged in the uniform practice of requesting unnecessary documentation, seeking repetitive submissions of the same documents, and never rendering a decision, and in doing so, such activity provides the basis for an unfair inquiry and could deceive the public because it places burdensome requirements on me that served no purpose but to provide a basis for either engaging in Dual Tracking by selling my home and by denying loss mitigation options, such as loan modifications ; They failed to properly consider loan modification applications before trying to sell my home. \n\n1. On information and belief, Mr. Cooper does not own a significant majority of the loans on which it functions as servicer. \n\n2. Economic Factors that discourage Mr. Cooper from meeting its obligations under HAMP by facilitating loan modifications include the following : a. Mr. Cooper may be required to repurchase loans from the investor in order to permanently modify the loan. This presents a substantial cost and loss of revenue that can be avoided by keeping the loan in a state of temporary modification or lingering default. \n\nb. The monthly service fee that Mr. Cooper, as the servicer, collects as to each loan it services in a pool of loans, is calculated as a fixed percentage of the unpaid principal balance of the loans in the pool. Consequently, the modifying of a loan to reduce the principal balance results in a lower monthly fee to the servicer. \n\nc. Fees that Mr. Cooper charges borrowers that are in default constitute a significant source of revenue to it. Aside from income Mr. Cooper directly receives, late fees and process management fees are often added to the principal loan amount thereby increasing the unpaid balance in a pool of loans and increasing the amount of the servicers monthly service fee. \n\nd. Entering into a permanent modification will often delay a servicers ability to recover advances it is required to make to investors of the unpaid principal and interest payment of a non-performing loan. The servicers right to recover expenses from an investor in a loan modification, rather than a foreclosure, is often less clear and less generous. \n\ne. Fixed overhead costs involved in successfully performing loan modifications involve up-front costs to the servicer for additional staffing, physical infrastructure, and expenses such as property valuation, credit reports and financing costs. \n\n3. Rather than allocating adequate resources and working diligently to reduce the number of loans in danger of default by establishing permanent modifications, Mr. Cooper has serially strung out, delayed, and otherwise hindered the modification processes that it contractually undertook to facilitate when it accepted billions of dollars from the United States. Mr. Cooper delay and obstruction tactics have taken various forms with the common result that my loan serviced by Mr. Cooper, in which im eligible for permanent loan modifications, and have met the requirements for participation in HAMP, have not received permanent loan modifications to which iam entitled. \n\n4. By failing to comply with the California Homeowner Bill of Rights ; specifically, by engaging in the unlawful practice of Dual Tracking, Mr. Cooper has left me in a state of despair and panic as i seek a loan modification from Mr. Cooper, only to have my home wrongfully, oppressively and illegally sold out from under my feet. \n\nWe have been in a fighting battle with our bank/ servicer Mr. Cooper for a very long time i mean very long time. We have been trying for a loan mod. recently its getting worse and worse and the bank keep pressuring and not providing an affordable loan mod. that we can afford which will help us keep the home and not throw us out on the streets. So our servicer now is Mr. Cooper and we have sent them many times a request for mortgage assistance forms packet with all documents. Recently we mailed and faxed in a copy of the request for mortgage assistance and all required documentation to be reviewed for a loan modification. Mr. Cooper keeps prolonging this process. Mr Cooper now are saying that have all documents received for a review but they still have our home under a foreclosure sale and recently they said they put the sale on hold but different representatives keep telling us different information one says ins on hold the other says its still on sale but after reviewing and calling the trustee its still on for sale, seems to me some of the representatives are trying to manipulate and get me false information to think that everything is ok but its not according to the trustee nothing has been stopped or placed on hold. \n\nIf Mr. Cooper proceeds to sell my home while having a complete request for loan modification review packet then they will be in violation of multiple laws including but not limited to my California Home Owner Bill of Rights Act Mr. Cooper has still placed my home on a foreclosure sale set for XX/XX/XXXX and has not CANCELLED the foreclosure sales until the review and an affordable loan modification and the process has been approved and agreed upon not some {\$4000.00} monthly payment that no one can afford but one that is reasonable and that if helping the borrower not a ridiculous sum that will force us to move out. We sent them all documents and are in review at the moment but why have they not yet removed the foreclosure sale. \n\nThank you.'

Product: Mortgage

Text Summarization

Automatic text summarization is the task of producing a concise and fluent summary while preserving key information content and overall meaning

Two different approaches

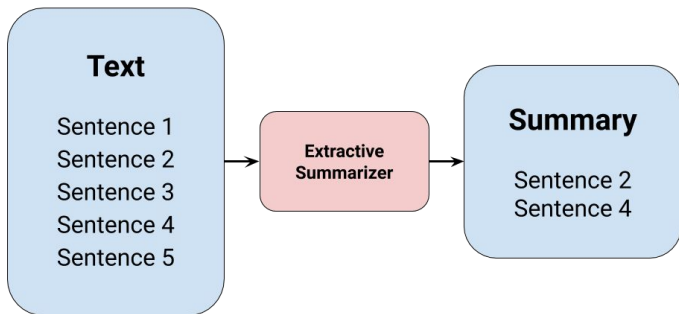


<https://openai.com/blog/fine-tuning-gpt-2/>

Summarization

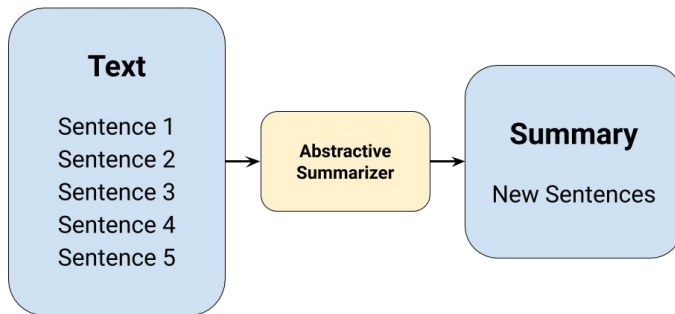
Extractive Summarization

- Identify the important sentences or phrases from the original texts
- Extract **only** sentences from the text
- Extracted sentences will be our summary



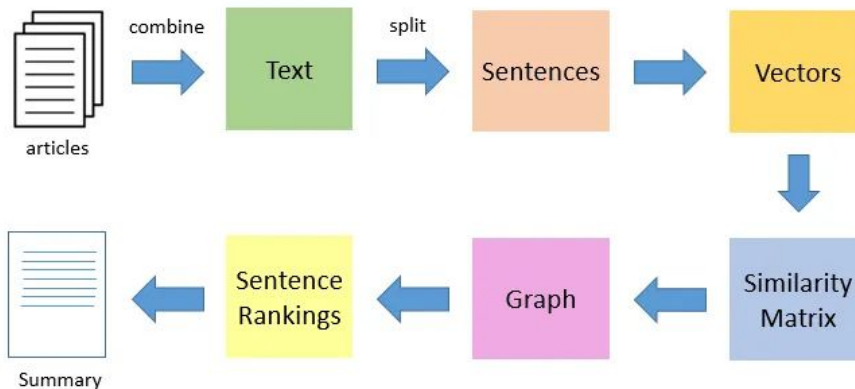
Abstractive Summarization

- Generate new sentences from the original text
- Generated sentences may not be present in the original text
- Much more difficult task involving language



Extractive Summarization: TextRank Algorithm

- **Extractive** and **unsupervised** text summarization technique
- Fairly old - proposed in 2004
<https://www.aclweb.org/anthology/W04-3252>
- Similarities between TextRank and PageRank:
 - Sentences in place of web pages
 - Similarity between any two sentences used as an equivalent to the web page transition probability
 - Similarity scores stored in a square matrix, similar to matrix M used for PageRank

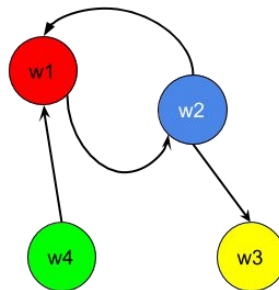


Understanding PageRank

- Google's **PageRank algorithm**, used primarily for ranking web pages in online search results
- Outputs a probability distribution representing the likelihood that a person randomly clicking on links will arrive at any particular page.
- Assign weights for web pages by the following formula:

$$S(V_i) = (1 - d) + d * \sum_{j \in In(v_i)} \frac{1}{|Out(V_j)|} S(V_j)$$

- $S(V_i)$ - weight of webpage i
- d - damping factor, for case of no outgoing links
- $In(V_i)$ - inbound links of i
- $Out(V_j)$ - outgoing links of j
- $|Out(V_j)|$ - the number of outbound links



Transition Matrix **M**

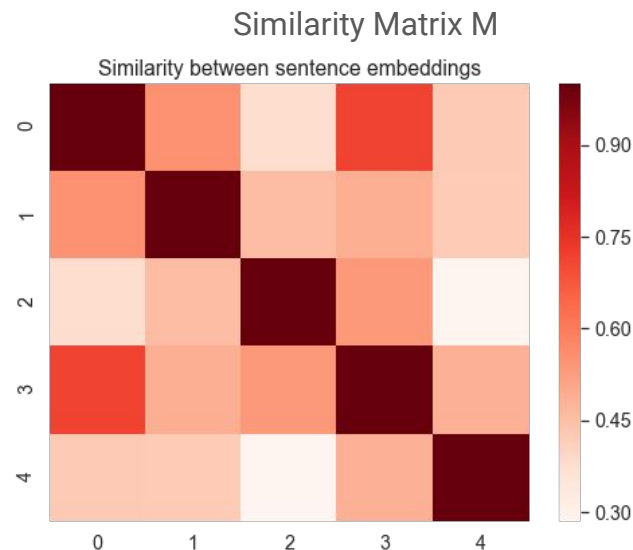
	w1	w2	w3	w4
w1	0	0.5	0	0.5
w2	0.5	0	0.5	0
w3	0.25	0.25	0.25	0.25
w4	1	0	0	0

- Deducing the importance of the web pages from global information recursively drawn from the whole graph

TextRank and Similarity Matrix

- Same idea - compute similarity between each pair of sentences. This build up a complete graph, i.e. each vertex (sentence) is connect to every other one, and each edge is the similarity score between two vertices
- Value of the edges basically like 'casting a vote' for the other vertex - higher votes results in higher importance for the vertex (sentence)
- But who voted also matters → Just like in PageRank, important sentences have a higher contribution when voting
- In the old days they calculated similarity between sentences using overlaps, now we can use embeddings

- 'I have reached out to the credit bureaus about these debts they have put on my credit file.',
- 'This has been a long road.',
- 'When ever I apply for anything I am never approved because for some reason I believe me and someone has the same social.',
- 'Its extremely frustrating.',
- 'I just would like all this inaccurate information removed'



Summary Extraction

'Around the middle of XX/XX/XXXX I received a call from a man named XXXX from a company called True Solution Inc. XXXX XXXX XXXX XXXX. XXXX informed me that this company was working on behalf of the Department of Education to assist me with consolidating my student loans and getting me into a reduced payment plan based on my income/dependence. He walked me through what my payments would be, and had me sign the contract via email. I contacted the Department of Education and my current loan holder (XXXX) and both said they did not know anything about this company. The Department of Ed told me to contact the Federal Trade Commission and the XX/XX/XXXX. In the mean time I blocked my credit card information. I received a letter from a company called XXXX XXXX XXXX saying my loan had been transferred to them as of XX/XX/XXXX, with a balance owed of {\$1200.00}. I called them and they referred me back to True Solution Inc. when I called them back I was told their was nothing I could do about it, I signed a contract and that an attorney would do nothing to help me. They said they would report this to my credit report. They claimed that they are charging for the work of consolidating my loans. The manager that I spoke to laughter in the phone, told me I wasnt competent enough to talk to him and that I needed to have my attorney contact him. He gave gave me his name as XXXX and his number as XXXX. I spoke with the Department of Education and they stated that consolidating my loans was a free service through them and that setting me up on a percentage repayment plan was a free service from them as well and that I should not have To pay anyone to do this. I have not paid anything to either of the companys but need this removed from my credit. I was informed that I was scammed. The XXXX XXXX XXXX agreed. I still going to reach out to other organizations before contacting my attorney to help resolve this. The company has only been in service less than a year and has a few complaints listed already. I have called and asked them to send me a copy of my contract and still have not received it. It was supposed to be sent to me via email.'



'I received a letter from a company called saying my loan had been transferred to them as of with a balance owed of {\$1200.00}. informed me that this company was working on behalf of the Department of Education to assist me with consolidating my student loans and getting me into a reduced payment plan based on my income/dependence. I contacted the Department of Education and my current loan holder and both said they did not know anything about this company. I have called and asked them to send me a copy of my contract and still have not received it. I spoke with the Department of Education and they stated that consolidating my loans was a free service through them and that setting me up on a percentage repayment plan was a free service from them as well and that I should not have To pay anyone to do this.'

Product: Debt collection

Issue: Took or threatened to take negative or legal action

Summary Extraction

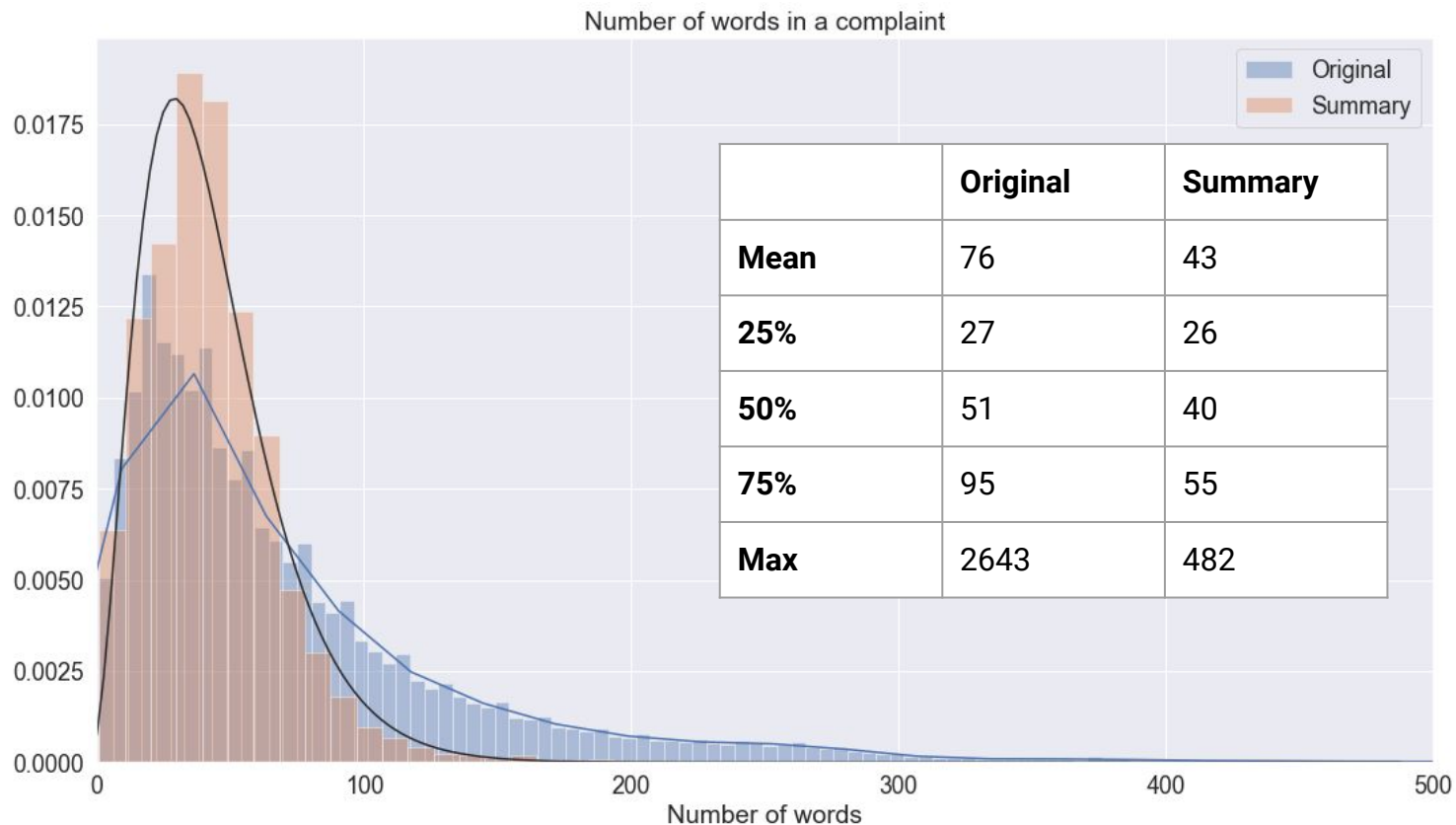
'It was recently reported on my account that I had a derogatory account with Eastern Connecticut Account. I have never received anything regarding a missing payment to my New York address. This is regarding an XXXX XXXX payment for XXXX dollars. This was a charge from XX/XX/XXXX, I had XXXX XXXX XXXX XXXX at the time as I was under my father's insurance. \tI received a few phone calls in XX/XX/XXXX from *XXXX Fraud Alert #*, I obviously did not answer. I had a voicemail saying to please call and confirm a social security number. I XXXX Eastern Connecticut and saw extremely negative reviews. My concern regarding giving them my personal information only began to grow. When I called back I instructed the representative that they showed up on my phone as fraud. Then the representative began to instruct me on an XXXX XXXX debt. I asked from when was this from, they said XX/XX/XXXX. I asked her if XXXX XXXX put my insurance in. XXXX XXXX XXXX XXXX XXXX in and out of state number. I know that my insurance would have went through, as I had XXXX XXXX XXXX and had numerous bills that were all covered. I instructed her to reach out to the XXXX XXXX and try to submit the claim or call me back if there were any issues. I didn't feel comfortable just making a payment to this organization, especially after the reviews and the type of infection I had during that time. When the organization never called me back, I assumed that the XXXX XXXX had taken care of it.\tAfter I got an update from my XXXX app on a negative report to my score, I saw it was Eastern Connecticut. I immediately called Eastern Connecticut, and left numerous Voice Mail Messages. I had one representative hang up on me multiple times, when I was trying to explain I don't live in Connecticut. After finally getting through to one representative, I finally paid the XXXX dollars. I asked if they could send a report to your company saying they didn't have my address and the person laughed at me over the phone. He said, "Why would I help your credit score". He said, having a phone number is sufficient enough to report someone as a derogatory account. \tThis organization is taking advantage of people and I don't want my credit report to be negatively reflected by this. XXXX even reported their entity as fraud. I did pay this organization, however extremely reluctantly. I have attached a copy of the receipt. Please, if your organization can update my credit score and remove this derogatory account information. I have never missed a payment on any of my accounts. I have a number of student loans that I'm trying to pay off, and this negative credit score will hurt my future in receiving a lower interest rate.'



'He said, having a phone number is sufficient enough to report someone as a derogatory account. This organization is taking advantage of people and I don't want my credit report to be negatively reflected by this. I asked if they could send a report to your company saying they didn't have my address and the person laughed at me over the phone. I have never received anything regarding a missing payment to my New York address. I have a number of student loans that I'm trying to pay off, and this negative credit score will hurt my future in receiving a lower interest rate.'

Product: Debt collection
Issue: Written notification about debt

Summary Extraction



Back to Product Classification

- Use exactly the same models as before but with only 5-sentence summary
- Logistic regression: accuracy = 0.78
- Linear SVC: accuracy = 0.80
- Very little loss in accuracy compared to using the full text.

Scaled recall matrix

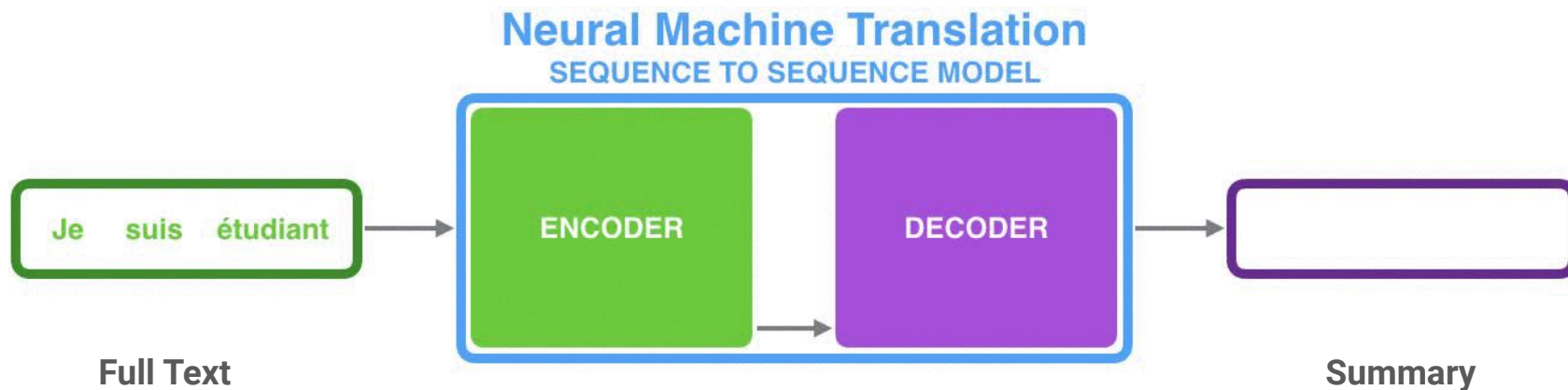
Banking Service	0.78	0.02	0.05	0.00	0.00	0.31	0.01	0.08	0.00
Consumer Loan	0.01	0.66	0.01	0.01	0.01	0.02	0.01	0.30	0.01
Credit Card	0.10	0.06	0.77	0.01	0.02	0.19	0.00	0.11	0.01
Credit Reporting	0.05	0.41	0.15	0.75	0.13	0.06	0.06	0.30	0.14
Debt Collection	0.03	0.17	0.06	0.05	0.77	0.09	0.03	0.61	0.09
Money Transfer	0.03	0.00	0.01	0.00	0.00	0.81	0.00	0.02	0.00
Mortgage	0.02	0.06	0.01	0.01	0.01	0.06	0.89	0.19	0.03
Payday Loan	0.01	0.05	0.01	0.00	0.00	0.01	0.00	0.66	0.01
Student Loan	0.00	0.02	0.00	0.00	0.01	0.01	0.00	0.10	0.88
	Banking Service	Consumer Loan	Credit Card	Credit Reporting	Debt Collection	Money Transfer	Mortgage	Payday Loan	Student Loan

What's next?

Abstractive summarisation and deep learning

Abstractive Summarization

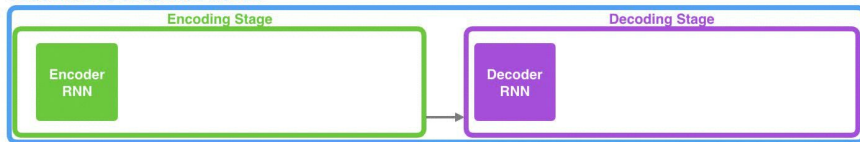
- A lot harder to do!
- Deep Learning
- Sequence-to-Sequence (Seq2Seq) Modelling - Encoder-Decoder structure



Recurrent Neural Networks and Attention

Neural Machine Translation

SEQUENCE TO SEQUENCE MODEL



A unrolled view of the sequence-to-sequence encoder-decoder model for NMT, where both encoder and decoders are RNNs with either GRU or LSTM units

Attention:

The encoder passes all the hidden states (associated with each word) from the encoder to the decoder

The decoder assigns a score to each hidden state, and use softmax to amplify hidden states with high scores and drowning out the states with low scores

Enhancements

- Multi-layer bidirectional RNNs
 - Transformers
 - Pointer generator
 - Reinforcement learning with Seq2Seq
-
- Recent development: Fine tune a pre-trained GPT-2 model from Open AI with **774 million** parameters with human labellers - <https://openai.com/blog/fine-tuning-gpt-2/>
-
- Interesting to see how they deal with verbose complaints!

References

- TextRank: <https://www.aclweb.org/anthology/W04-3252>
- GloVe: <http://www.aclweb.org/anthology/D14-1162>
- Jay Alammar's beautiful series of blogs on Seq2Seq and Transformers
<https://jalammar.github.io/illustrated-transformer/>
- Build your own summarizer guide <https://hackernoon.com/text-summarizer-using-deep-learning-made-easy-490880df6cd>

mudano.com
info@mudano.com



@everMudano