Consumer Financial Protection Bureau

Complaint Narrative Summarization

Cora Li

Main Purpose

- Users: Workers who process complaint narratives
- Benefit: Summarization can help them grasp important information and consequently improve processing effiency

Consumer Financial Protection Bureau

Two Approaches

- Text Summarization

-Extract important sentence in narratives

- Sub-issue Prediction

-Sub-issue can be blank

-Sub-issue variable gives detailed summary of problems than issue variable

Text Summarization

- Algorithm:

-Assume frequent word is important
-Calculate word frequency in narratives (except for stop words and punctuation)
-Extract sentence containing frequent terms

- Benefit:

-Achieve summarization without other input
-No black box

- Disadvantage:

-Wrong assumption
-Extraction miss important information

Consumer Financial Protection Bureau

Sub-issue Prediction

- Model:

-Logit Model

-SVM model

-Pre-trained word embeddings model

-LSTM model

- Benefit:

-Advanced method

-Sub issue returns summary instead of extraction

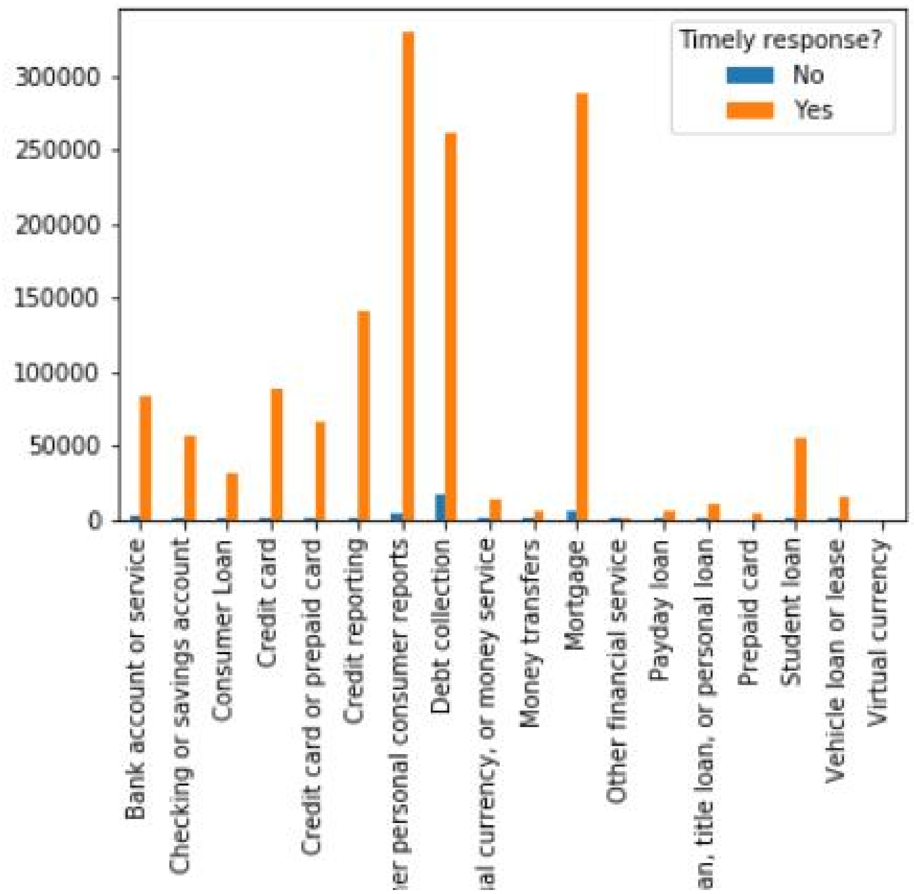
- Disadvantage:

-Need additional input data (sub-issue variable)

-Too many sub-issue as dependent variable: Take subset of narratives with similar product/issue

Sub-issue Prediction - Dataset

- Overview of data with narratives



- **Focus on**
- **Products of loan**

-Payday loan, title loan, or personal loan

-Credit reporting, credit repair services, or other personal consumer reports

-Payday loan

- **Issue: Incorrect information on your report**
- **Total 75799 narratives**

Sub-issue Prediction-Model performance

- Logit model

	precision	recall	f1-score	support
Information belongs to someone else	0.76	0.77	0.77	1939
Account status incorrect	0.80	0.78	0.79	1935
Account information incorrect	0.84	0.83	0.84	1851
Personal information incorrect	0.96	0.97	0.97	1965
Old information reappears or never goes away	0.94	0.92	0.93	1978
Public record information inaccurate	0.96	0.97	0.97	1906
Information is missing that should be on the report	0.95	0.97	0.96	2019
accuracy			0.89	13593
macro avg	0.89	0.89	0.89	13593
weighted avg	0.89	0.89	0.89	13593

- SVM model

	precision	recall	f1-score	support
Information belongs to someone else	0.84	0.85	0.84	1939
Account status incorrect	0.87	0.84	0.86	1935
Account information incorrect	0.87	0.83	0.85	1851
Personal information incorrect	0.97	0.99	0.98	1965
Old information reappears or never goes away	0.95	0.95	0.95	1978
Public record information inaccurate	0.97	0.98	0.97	1906
Information is missing that should be on the report	0.97	0.98	0.97	2019
accuracy			0.92	13593
macro avg	0.92	0.92	0.92	13593
weighted avg	0.92	0.92	0.92	13593

Sub-issue Prediction-Model performance

- Pre-trained word embeddings model

	precision	recall	f1-score	support
Information belongs to someone else	0.46	0.01	0.02	1070
Account status incorrect	0.62	0.18	0.28	1107
Account information incorrect	0.94	0.46	0.62	2444
Personal information incorrect	0.06	0.85	0.11	135
Old information reappears or never goes away	0.36	0.34	0.35	321
Public record information inaccurate	0.75	0.67	0.71	323
Information is missing that should be on the report	0.82	0.45	0.58	263
micro avg	0.45	0.34	0.39	5663
macro avg	0.57	0.42	0.38	5663
weighted avg	0.72	0.34	0.42	5663
samples avg	0.34	0.34	0.34	5663

- LSTM model

	precision	recall	f1-score	support
Information belongs to someone else	0.76	0.17	0.28	1070
Account status incorrect	0.64	0.41	0.50	1107
Account information incorrect	0.91	0.69	0.79	2444
Personal information incorrect	0.28	0.47	0.35	135
Old information reappears or never goes away	0.48	0.48	0.48	321
Public record information inaccurate	0.70	0.82	0.76	323
Information is missing that should be on the report	0.61	0.70	0.65	263
micro avg	0.74	0.53	0.62	5663
macro avg	0.63	0.54	0.54	5663
weighted avg	0.76	0.53	0.60	5663
samples avg	0.53	0.53	0.53	5663

Consumer Financial
Protection Bureau

Thank You~