Identifying complaints on social media

Using Federated Learning

Github Link

Mentor:
Dr. Sriparna Saha
Associate Professor
IIT Patna

By:
Priyanka Sachan
1901CS43

What is a complaint?

"A complaint presents a state of affairs which breaches the writer's favorable expectation ."

(Olshtain & Weinbach, 1987)

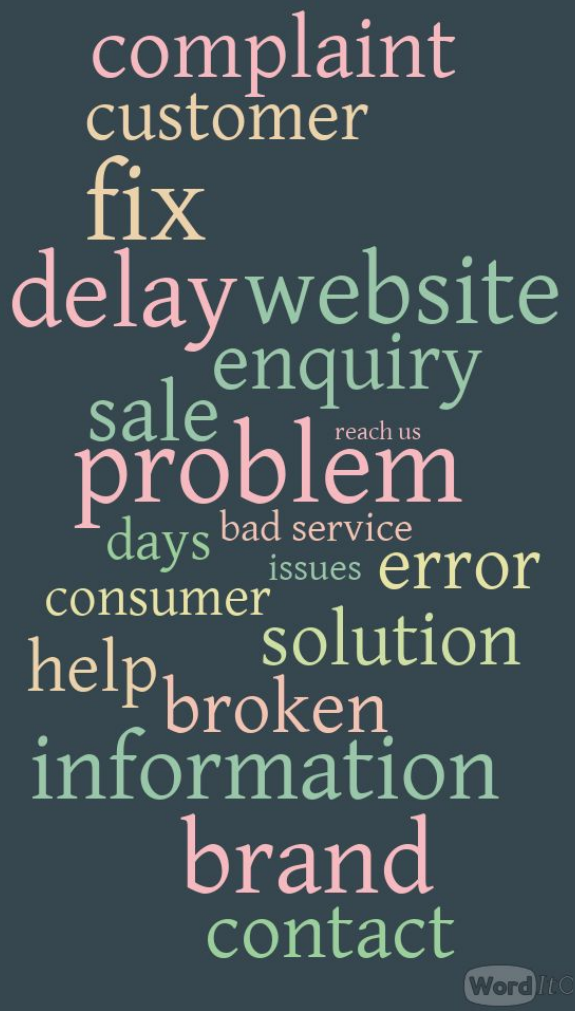
Example

Complaint:

"@Topman what's the point in paying for express delivery if you can't keep to it? Waste of time."

Non-Complaint:

"@BN_care will i still receive it in the 2-3 express delivery time frame?"



Motivation

- Improve customer experience by addressing client concerns
- Identify complaining intent in dialogues in NLP applications
- Understanding context and types in language
- Understanding human traits specific of complaining
- All while keeping the data privacy intact...

Past Work

- Feature-based and task specific machine learning models *(Preotiuc-Pietro et al., 2019)*
- Transformer based model *(Mali Jin, Nikolaos Aletras 2020)*

Federated Learning

DataSet Used (from twitter)

https://github.com/danielpreotiuc/complaints-social-media

3449 lines (3449 sloc) | 423 KB

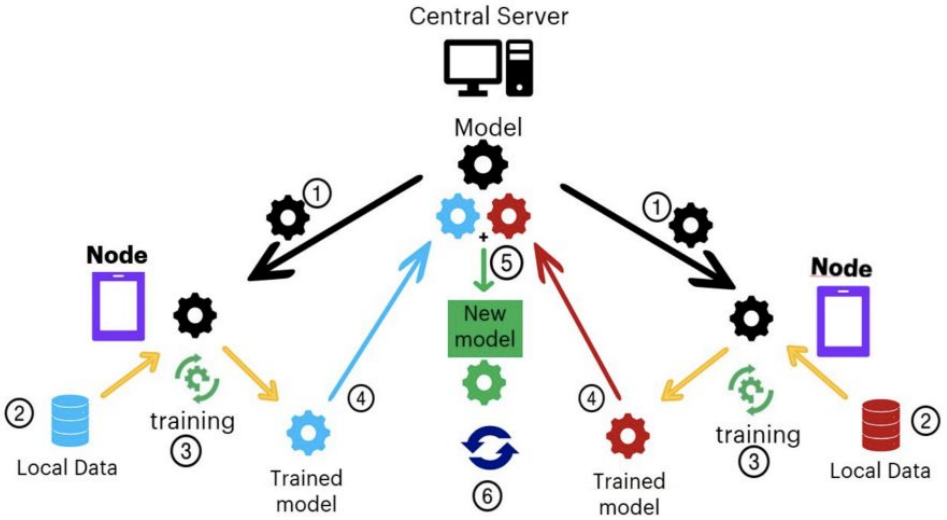
Raw | Blame

Search this file...

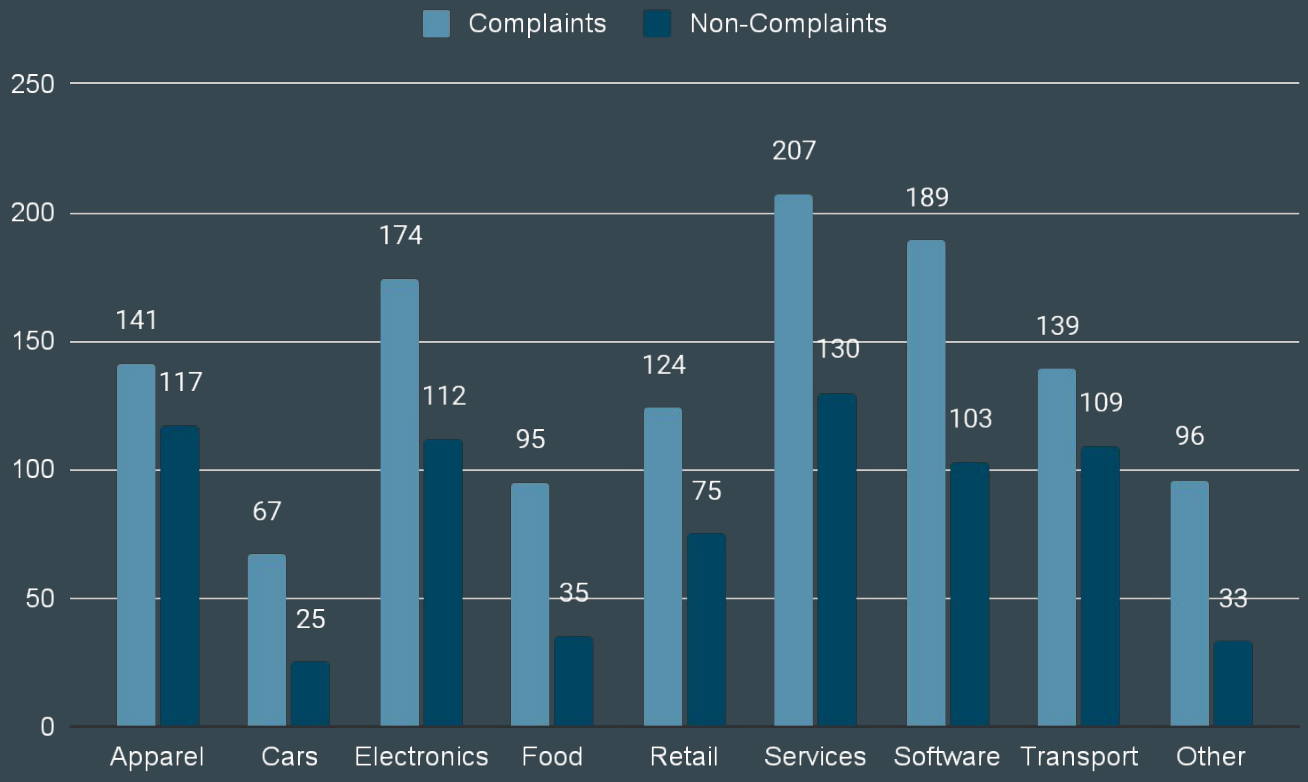
1	**22631081115127808**	**@FC_HELP can I return online purchases to a House of Fraser concesssion if I have the delivery note and order details?**
2	23364081385734144	@FC_Help Hi - I'm writing a piece for MSN Him and wondered who I could talk in the PR dept about this season's clothes/trends, etc? Help!
3	25550410899005441	@FC_Help i need to check my order
4	29494955818876928	@FC_Help I need to get in contact with someone regarding the fc.com website. I work for a major search engine. Tried enquiries@ and web form
5	32523695972945920	@FC_Help How can I get a hold of you so we can discuss the problem I am havnig with my coat?
6	34600441576824832	@FC_Help Will you be getting the wendy cotton v neck dress in pavlova back in stock on the site?
7	48757563688353792	Build new system? @nvidia forgot to put a bootable USB stick with the cards. So buy another card, install with that and put the fermi in.
8	49340019671973888	@NVIDIAGeForce If I connect a #GTX580 to a TV through HDMI will I get audio output as well?
9	49911462155665409	@FC_Help Please could you point me in the direction of your PR department. I need to contact them with regards to some client work.
10	51182036098093056	@FC_Help looking for models / actors or dancers ? Olivier
11	61380449364230144	@FC_Help hi m order is 913181 did you revise the money? if you did.. how about the shipping ?
12	70533385910620160	@FC_Help looking for "bright carol" or "stained glass" dress. do you have these in stock anymore?
13	70859249336598528	@FC_Help not sure if those are official names for dress, but here is image of it. I really want to find it somewhere! http://bit.ly/iMCQsl
14	70879180652412931	@FC_Help Thanks! Do you have any idea of the actual name of the dress?
15	79715568042770433	@nvidiacc Anyone know what's up with the Geforce 500 series 580 GPX Driver 275.33
16	83562837473103872	Dear @nvidia, I don't think I should have to roll back to driver v270.61 to make my games work, and my desktop not glitch out.

Federated Learning

Federated learning enables multiple actors to build a common, robust machine learning model without sharing data, thus allowing to address critical issues such as data privacy, data security, data access rights and access to heterogeneous data.



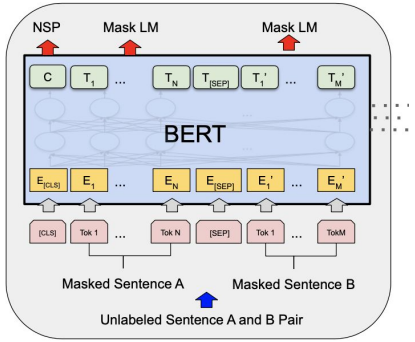
It includes 1,232 complaints (62.4%) and 739 non-complaints (37.6%) over 9 domains (e.g. food, technology, etc.).



Architecture

Best Model from Mali Jin, Nikolaos Aletras paper (Dec 2020)

BERT



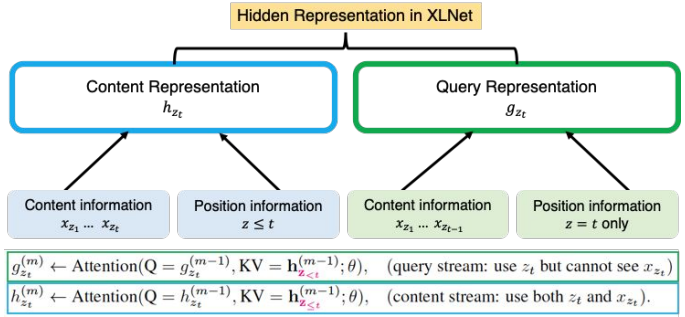
Pre-training

XLNet Models

XLNet + Regression Layer

XLNet + LSTM

XLNet + CNN



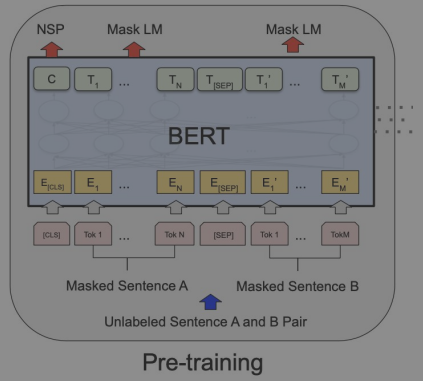
Comparison between models

Model	Accuracy	Precision	Recall	F1 Score
Bert (from paper)	88.0 ± .03	87.1 ± .03	87.3 ± .03	87.0 ± .03
XLNet Sequence Classification (XLNet + Regression Layer)	89.13 ± .18	83.84 ± .39	86.49 ± .28	85.06 ± .20
XLNet Embedding + LSTM	72.11 ± 3.4	63.87 ± 7.4	50.98 ± 6.7	56 ± 5.8
XLNet Embedding + CNN	82.89 ± 2.0	83.45 ± 2.3	65.32 ± 4.6	73.15 ± 2.5

Architecture

Best Model from Mali Jin, Nikolaos Aletras paper(Dec 2020)

BERT

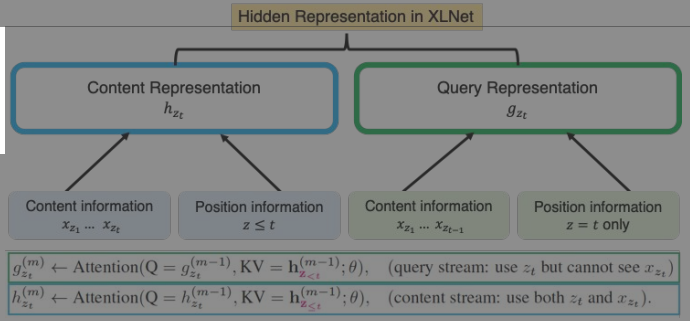


XLNet Models

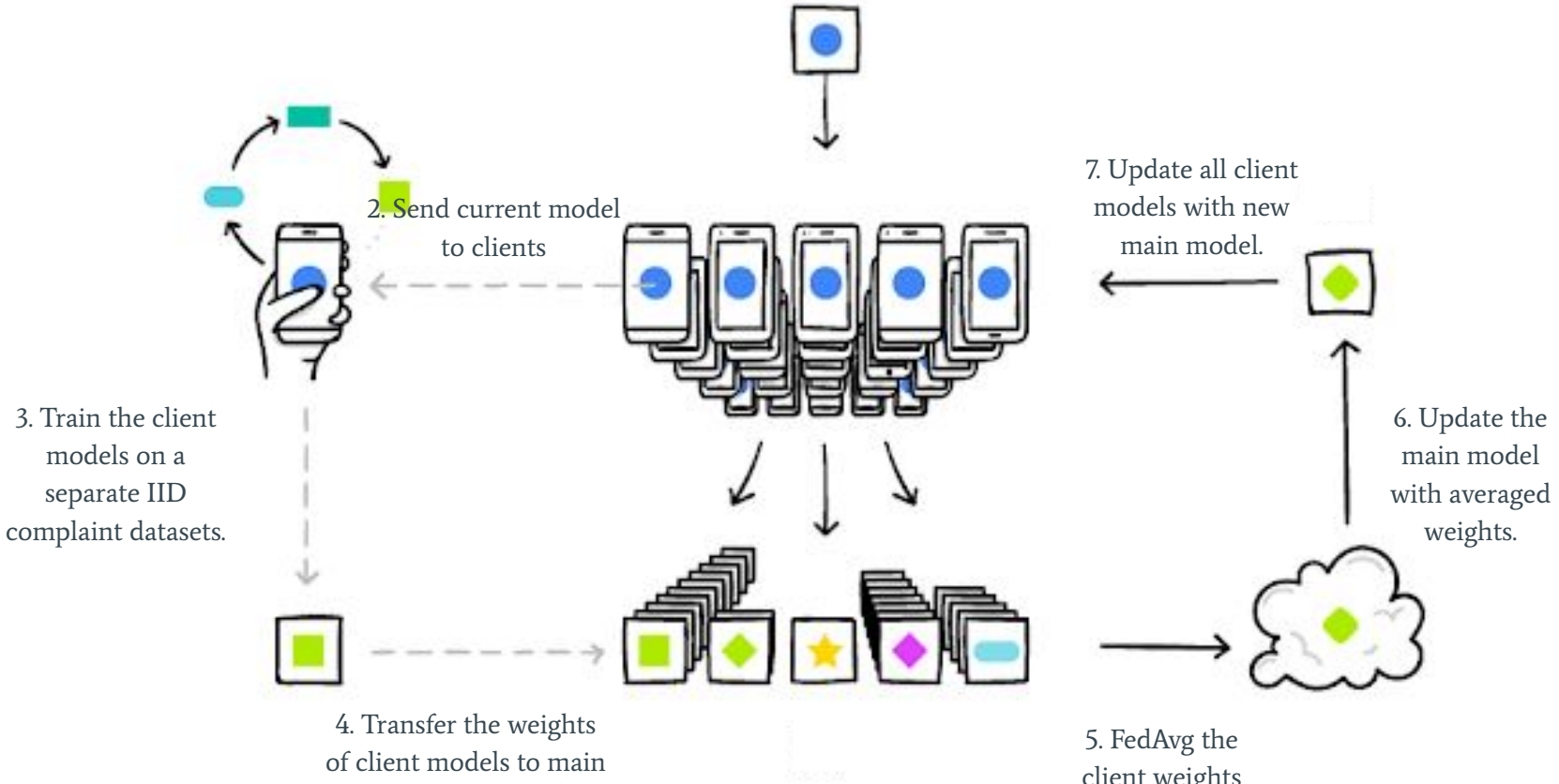
XLNet + Regression Layer

XLNet + LSTM

XLNet + CNN



Workflow


1. Train pretrained xLNet main model on Amazon Reviews dataset (Transfer Learning).
2. Send current model to clients
3. Train the client models on a separate IID complaint datasets.
4. Transfer the weights of client models to main model
5. FedAvg the client weights
6. Update the main model with averaged weights.
7. Update all client models with new main model.

Results of Federated Averaging over 4 nodes over IID Dataset:

Model	Accuracy	Precision	Recall	F1
Main Model (Before FA) Trained on Amazon Reviews	75.94	64.67	86.86	74.14
Main Model (After FA)	**85.50**	**80.41**	**83.94**	**82.14**
Client 1	84.34	77.12	86.13	81.37
Client 2	87.25	82.51	86.13	84.28
Client 3	84.30	79.02	82.48	80.71
Client 4	87.24	84.96	82.48	83.70

Error Analysis

Implicit

“It started yesterday , but I try again it could work normal. But since last night its just like this .”

Rarely include words related to complaints (e.g. ‘disappointed’, ‘bad service’)

Irony

“Thank you so much for making a box that shreds apart even when carried by both handles.”

Using terms such as ‘congratulations’, ‘thank you’ and ‘brilliant’.

Interrogative tone

“Folks , what is cost of text message to a us number?”

Contains negative sentiment

“This would be a terrible idea https://t.co/bqoUYsSSQv”

N-IID EXPERIMENT 1: Train local domain models, fedAvg it, then test local and main models across all domain tweets.

Model	Accuracy	Precision	Recall	F1
Main Model (Before FA) **Trained on Amazon Reviews**	76.61	80.95	81.6	81.27
Main Model (After FA)	**81.09**	**82.70**	**88.0**	**85.27**
Apparel	83.58	90.35	82.4	86.19
Cars	79.10	80.29	88.0	83.96
Electronics	83.58	89.65	83.2	86.30
Food	77.11	76.87	90.4	83.08
Retail	78.10	79.56	87.2	83.20
Services	82.08	88.03	82.4	85.12
Software	83.08	85.82	87.2	86.50
Transport	83.58	88.33	84.8	86.53
Other	78.60	79.71	88.0	83.65

N-IID EXPERIMENT 2: Train local domain models over some domains (eg. 0,1, 2, 3,4), fedAvg it, then test main models across all other (eg. 5,6,7,8) domain tweets.

Model	Accuracy	Precision	Recall	F1
Main Model (Before FA) **Trained on Amazon Reviews**	76.61	80.95	81.6	81.27
Trained on (Apparel+Cars+Electronics+Food+Retail) and Tested on (Services+Software+Transport+Other)	83.64	84.62	90.28	87.36
Trained on (Services+Software+Transport+Other) and Tested on (Apparel+Cars+Electronics+Food+Retail)	84.03	87.23	87.08	87.15

Takeaways

- Complaints can be predicted with good accuracy & data privacy .
- Transformer models struggle with implicit complaints and sarcasm .
- Transformer models perform robustly for domain pairs where these domains are used for either training or testing .

References

- Preotiuc-Pietro, D., Gaman, M. and Aletras, N., 2019. Automatically identifying complaints in social media. arXiv preprint arXiv:1906.03890.
- Mali Jin, Nikolaos Aletras . 2020. Complaint Identification in Social Media with Transformer Networks, The 28th International Conference on Computational Linguistics, Underline Science Inc.
- Qiang Yang, Yang Liu, Tianjian Chen, Yongxin Tong. 2019. Federated Machine Learning: Concept and Applications. arXiv preprint arXiv:1902.04885
- Agrin Hilmkil, Sebastian Callh, Matteo Barbieri, Leon René Sütfeld, Edvin Listo Zec, Olof Mogren. 2021. Scaling Federated Learning for Fine-tuning of Large Language Models. arXiv:2102.00875

Thank You.