Project Plan

Team -13 (ICT Dept)

Ishan Mistry - 18BIT033
Sagar Sinha - 18BCP096
Sanskar Bhuwania - 18BCP099
Anurag Pal - 18BCP007

Under the Supervision and Guidance of
Dr. Santosh Kumar Bharti
Department of Computer Science and Engineering
Pandit Deendayal Energy University

Dr. Pawaan Sharma
Department of Information and Communication Technologies
Pandit Deendayal Energy University

19CPIC01
Comprehensive Project

Title

"Identification of the critical point in a conversation and use it to identify and prevent cancel culture using NLP techniques"

Objectives

Modeling the entire conversations in a computationally efficient manner, preserving the various relationships(both intrinsic and extrinsic) between the sentences of the conversation. the so that we can later on perform analysis on the various kinds of conversations and henceforth use these observed patterns to train a generative model.
Once we have a set framework of how to store these sentences we can have our own custom metrics and study all the various categories of conversations - Conflicts/collaboration/mental health related, etc. and try to encode them into a single metric that we would aim to optimize.This final metric will take into consideration all the features(Relevance to the topic, history, sentiment etc.) that will be extracted from the EDA of the conversations.
This critical point in a conversation can help us pinpoint the causes(sentences) of contention and groupthink, and thereafter predict a sentence that could moderate the discussion which otherwise would have turned into a heated argument or herd thinking."

Motivation and Applications

It happens everywhere around social media conversations that just because there are few differences in how people converse and present ideas differently there's a high chance of some sort of confusion or disagreement even when the topic is something that both of them have in common. This is almost never a sudden jump from Strongly agree to Strongly Disagree but rather there are many instances/points sprinkled throughout the conversation that leads the other side to think that they are against them.The aim of the model will be to identify these clues and help mitigate such situations proactively.

If left unattended, such conversation could become lopsided easily and people with differing views might become the centre of attack as people try to vehmently present their opinions.Such conversations are practically dead ends when they could have been fruitful if moderated properly. Cancel culture or call-out culture is a modern form of ostracism in which someone is thrust out of social or professional circles – whether it be online, on social media, or in person. This is getting out of bounds these days because the anonymity of the individuals gives them this extraneous power over others online. We do realize that such events can have its own consequences and do surely take a toll on the victim.

One the key features of this project is to identify and quantify the latent patterns in the conversational data. This study could be a general eye-opener which will help us quantify the internal biases (confirmation bias etc) that we undergo subconsciously

and hence help us better understand how we communicate and whether we are victims to such biases.

This is a general idea and hence with little tweaks could be transferred to other domains as well. For example a model trained to capture conflicts can later on be used to identify sentences that are pretty vague and susceptible to be misunderstood.

The critical point can be viewed very differently depending on the context of the problem at hand:

1. Points of collaboration in a conversation
2. Resolving conflicts
3. Facilitating discussions that are vague
4. Steering the conversations - amplify the seemingly weaker ideas
5. Preventing cancel culture
6. Preventing Groupthink
7. Identifying suicidal tendencies

A system that can be used to transfer its learnings from one domain to another would be interesting to observe - maybe some new insights that we did not know earlier could be brought to light.

Tentative Workflow

The project is somewhat of a concoction of many projects in itself, so establishing meaningful milestones would be essential :

Phase 1:

1. Coming up with either “the best existing representation for modeling conversations” or “a novel representation/compression technique to capture all the relationships for sentences in a conversation”.
2. Testing the robustness and efficacy of these new representations!

Phase 2 :

1. Using this new representations to train a generative model that would be able to predict better sentences
2. Test this model against the SOTA models in the domain
3. Preparing a benchmark for the task and publishing the results

Tentative Workflow		
No.	Task	Objectives
1	Scope and Definition	- Feasibility - Areas to cover
2	Literature Review	- Conventional Techniques - NLP based models - Linguistical Modelling - Analysis for such similar tasks - Benchmarks - Evaluation
3	Datasets	New/Existing Datasets: - Where used? - Biased towards certain types? - Pre-processing done?
4	Previous Methods	- Implement/Run/Experiment - Test for our case - Limitations- Reasons -if not doing well?
5	Proposed Model/Representation	- Description - How do we try to capture everything in the talk? - Modelling the conversations/Compressing it - Rationale of building it up this way - How do we tackle these limitations?
6	Model Implementation and Training	- Data Gathering - Data Cleaning - Distributed Representations - Explore - Working on various spaces - individually - How do we unify these models? Or do we work on one model and try to train it to learn various ranges for choosing?
7	Debugging	Debug the model by analyzing where exactly the model goes wrong? Any patterns? - New features
8	Training again	Making these updates and training again
9	Validation	Validation and measuring performance
10	Benchmark	Ideate/Formulate a human Benchmark : - Form or something interesting - How do you filter out the noise? - Choose ques that can produce the best bang for the buck i.e shorter convos yet meaningful - Probabilistic analysis of the results we have - How much of it is by chance? Can you trust this human-based data?
11	Conclusions	Consolidate everything and start with completing the draft
12	Paper Submission	The work on this would run in parallel as it might be tough to recollect it later on. So, this phase is to be used to just put together the pieces collected so far.

Tentative Timeline :

Week 2 - Week 6	Literature Review
Week 7 - Week 9	Data Gathering and Cleaning
Week 10 - Week 12	Modeling the conversations
Week 12 - Week 14	Testing the efficacy of the representations
Week 15 - Week 16	Model Selection and Training
Week 17 - Week 19	Debugging + Training
Week 20 - Week 21	Validation + Testing + Comparison
Week 22 +	Submission and Conference prep