Multi-Target User Stance Discovery on Reddit

Benjamin Steel
School of Computer Science
McGill University
Montreal, Canada
benjamin.steel@mail.mcgill.ca

Derek Ruths
School of Computer Science
McGill University
Montreal, Canada
derek.ruths@mcgill.ca

Abstract

We consider how to credibly and reliably assess the opinions of individuals using their social media posts. To this end, this paper makes three contributions. First, we assemble a workflow and approach to applying modern natural language processing (NLP) methods to multi-target user stance detection in the wild. Second, we establish why the multi-target modeling of user stance is qualitatively more complicated than uni-target user-stance detection. Finally, we validate our method by showing how multi-dimensional measurement of user opinions not only reproduces known opinion polling results, but also enables the study of opinion dynamics at high levels of temporal and semantic resolution.

1 Introduction

People act in accordance with their opinions and beliefs (Bliuc et al., 2007). Therefore, efforts to understand and predict large scale human behaviour - from political opinion polls to consumer market studies - massively benefit from accurate maps of human opinions. There is ample evidence that social media is a valuable space in which to measure human opinions (Reveilhac et al., 2022). However, there are notable methodological gaps between the latest advances in natural language processing (NLP) on social media and the methods needed by practitioners (from analysts to social science researchers) who want opinion measurement that simply "works".

For practitioners, the ideal opinion measurement method has several properties: the ability to look at opinions of users on multiple topics at once, the ability to measure these at both high temporal and topical resolution, and to do this with minimal cost and time investment. In contrast, methodological research on this topic tends to focus on just one of these properties. For example, there is work that explores multiple topic measurement, but requires massive training datasets (Zhou et al., 2023). There are other studies that consider temporal and topical resolution, but do this only at the post, rather than user level (Li et al., 2021).

All this amounts to substantial progress in opinion measurement, but progress that is not directly useful to those who need to measure user opinions as the *starting point* of their studies.

We take the position that all the techniques to build such a practitioner-useful method actually exist - they need only be assembled. While this is happy news, we find that assembling these pieces together into a single, clear workflow is highly non-trivial and involves solving a highly non-trivial methodological issue. Namely, rendering the measurements taken of multiple user opinions comparable with one another. This becomes an issue because we may be much better at measuring opinion on one topic than on another. Nonetheless, we still want to ask questions about how these opinions interact with one another across user populations.

In this paper, we present a coherent workflow that assembles techniques from NLP at large into a recipe for the measurement of user stance in social media data at large scale. In the process, we show how to address the issue mentioned above, accurately assessing the opinions of a user on different topics with differing measurement error. We then validate our workflow by studying the multi-target opinions of users on parts of Reddit associated with Canada, showing that the resulting opinion data align with known trends, and that we can characterize previously unmeasured aspects of the online Canadian political discussion.

It is worth noting that this paper is not the final word on a practitioner-useful tool for multi-target opinion studies on social media. While our approach does assemble techniques into a clear, prescriptive workflow, manual tuning must still be done. Moreover, we have not provided a single software tool that automates the process. Both con-

Proceedings of the 14th Workshop on Computational Approaches to Subjectivity, Sentiment, & Social Media Analysis, pages 200–214
August 15, 2024

ceptually and practically, our recipe is hardly "plug-and-play". Nonetheless, we consider this an important step towards rendering the many advancements in opinion measurement useful for practitioners. Further, we submit our work as a starting point and indication of the need for future research that takes seriously the question of how to streamline multi-target user opinion measurement on social media.

We release the source code for our work at github.com/bendavidsteel/user-stance-discovery.

2 Background

Our pipeline has three stages: finding stance targets via topic modelling, classifing text stance via stance detection, and finally, inferring user stance via user stance detection. We will therefore review the background of each of these fields.

2.1 Topics Discovery

We define topic discovery here as the process of going from a raw text corpus of documents, and producing topical clusters of those documents.

Topic discovery has improved over purely bag of words clustering methods (e.g. latent dirichlet allocation (Blei et al., 2003)) with the use of language encoders and hierarchical clustering (Grootendorst, 2022), which aids with topic fidelity and tunable discovery. Multiple topic modelling steps has been shown to be effective for exploring political issues in a polarized Turkey (Rashed et al., 2021), a method of interest to us in its ability to discover viewpoints within a specific topic.

Beyond this, we would specifically like to be able to discover topics with heavily divided discussion, and for this methods for polarized topic discovery have surfaced (Paschalides et al., 2021), but have so far only been validated on news articles, and require a seed topic, and as such we will not be using this method here.

2.2 Text Stance Detection

Defined here as detecting the stance of a piece of text (e.g. favor, against, or neutral) towards a stance target (e.g. something one might have an opinion on, i.e. gun control).

Zero-shot stance detection is improving but not yet on-par with few-shot methods (Allaway and McKeown, 2023), showing that small amounts of training data are worthwhile for improving accuracy. Decoder-only language models (LLMs) are showing their utility for stance detection (Cruickshank and Ng, 2023), including using chain-of-thought methods (Zhang et al., 2023a), showing the effectiveness of prompt-based methods over classification head methods. Other work has provided extra data to the models, whether contextual data from social media, (Li et al., 2023) or additional descriptions of the stance target (Zhu et al., 2022). Additionally, work has shown stance detection benefits from multi-target-stance training (Li et al., 2021).

All these methods indicate the utility of few-shot data, prompting decoder language models, and additional context for improving text stance detection.

2.3 User Stance Detection

Here defined as inferring the stance of a user (a person expressing their stance on a platform through text posts) towards one or many stance targets.

Similar methods have been created for user stance detection previously, but many rely on platform specific features, where user stance views can be distinguished by following relevant accounts, or using specific, recognisable hashtags (Darwish et al., 2020; Samih and Darwish, 2021; Abeysinghe, 2023; Introne, 2023; Zhu et al., 2020; Jiang et al., 2023b; Zhang et al., 2023b). This technique can improve accuracy over text only features, but limits the method to only contexts where users interact with influential, opinionated extra-linguistic features, excluding other datasets without these features, or where this data is unavailable. It also ties the performance of the system to the extent that a stance of interest has associated prominent, opinionated entities.

Almadan et al. compare user-stance and text-stance opinion polling methods on Twitter, showing that user-stance produces more meaningful features for gauging public opinion (Almadan et al., 2023), but they only look at stance on vaccination, and use a pre-trained stance classifier trained on thousands of tweets. Zhou et al. focused on user-stance prediction of Weibo users (Zhou et al., 2023), but pre-selected stance targets, used more than 50,000 thousand labelled tweets for training, and focus on prediction as opposed to measurement. Wang et al. explore topical stance detection on an online discussion forum (Wang and Chen, 2021), however, they use sentiment analysis as a proxy for stance, which is poorly correlated with stance (Almadan et al., 2023). They also use likes, dislikes, and comment sentiment, as semi-supervision signals, which is poorly motivated, which, combined with dictionary based sentiment classification methods, results in poor accuracy. Kim et al. look at user stance on Reddit (Kim et al., 2023), but only look at one stance target, and do not control for the accuracy of their classifier on different labels.

With this work in mind, we will focus on building a user stance detection method that can work without platform specific features, use on the order of 10^1 training examples, and allows proper stance comparison by way of accounting for the accuracy of the classifiers used.

3 Pipeline Method

As currently practiced in the literature, user stance detection on a corpus of user-organized social media post data involves three stages (Almadan et al., 2023; Zhou et al., 2023; Wang and Chen, 2021):

1. **Stance target selection:** identifying the set of stance targets (i.e. something one might have an opinion on, e.g. gun control) on which we seek to measure each user's stance.
2. **Text stance inference on each stance target:** for each stance target, classifying each social media post in the corpus with a stance (e.g. favor, against, or neutral) on that target.
3. **User stance inference on each stance target:** for each user, aggregating the stance classifications for each of the user's posts to a user stance for that stance target.

Here we detail a pipeline that uses a host of already-existing methods to realize all three of these stages. Notably, because our aim is to measure the stance of each user on *multiple* targets at once, in Stage 3, we introduce a novel normalization approach to ensure that the user stance scores can be compared across targets.

3.1 Stance Target Selection

In this stage, our objective is to select the set of targets we will assess user stance towards. This is an intrinsically exploratory process. The practitioner will arrive at this step with an idea of the themes they want to study (e.g, "public health", "climate change", and "employment"). In this stage, we aim to distill these general themes into clearly defined stances that are both representative of the original intent and informed by the data available.

Our approach here uses an exploratory analysis of topics present within the data, focused by the themes we approach the study with.

We begin by running topic modelling on the post text (included titles and comments), to obtain a characterization of the most frequently discussed items in the corpus. To obtain these topics, we embed the texts with the sentence transformers model ‘all-MiniLM-L12-v2‘ (Reimers and Gurevych, 2019), reduced the dimensionality of the vector embeddings down to 5 dimensions using UMAP (McInnes et al., 2018), and clustered the reduced embeddings using HDBSCAN (McInnes and Healy, 2017), using the BERTopic library (Grootendorst, 2022).

At this point, we have topics, of which we manually select those we deem relevant to the original themes. For these selected topics, we seek to understand potential debates in order to find stance targets. We therefore take inspiration from Rashed et al. (2021), by further reducing the vector embeddings down to 2 dimensions, and re-clustering the data points, to find sub-topic discussion in each topic. These sub-topics provide a high-resolution picture of what themes are actually present in the data and in what relative abundance. Both presence and abundance are important to the selection of stance targets: it is impossible to measure user opinions about things that they have never mentioned.

The final step in this stage is for the practitioner to use the topic characterization, trends, statistics available to, combined with their own domain expertise, to define the stance targets for each topic themselves (i.e. choose the stance target *vaccine mandates* for a vaccines related topic).

3.2 Text Stance Detection

This stage focuses on inferring the stance of each post towards each of the stance targets selected in Stage 1. There are several steps involved here: (1) building training data, (2) training stance classifiers that work for each stance target, and (3) running the stance classifiers on the corpus posts.

Building training data. To train multiple stance detectors, we require training data for each in the form of annotated posts. Many stance detection systems can require thousands of training datapoints (Almadan et al., 2023; Zhou et al., 2023). From a practitioners perspective, it's important that we minimize the training dataset size to limit cost and coding time. As will be discussed, we chose a method where we found that coding 100 posts for each stance target, sampled from their respective topic, was sufficient: labelling them with their stance, with labels selected from ‘favor‘, ‘against‘, or ‘neutral‘. We used the definitions of stance from

Semeval-2016 (Mohammad et al., 2016). We used two annotators for each of these labelling tasks, ensuring that our annotations were sufficient quality by running an interannotator agreement statistic (Gwet, 2008). An adjudicator then chose the final gold stance label, by looking at the two annotator's labels.

Training stance classifiers. With labelled data in hand, we then train and test stance classifiers. There are multiple paradigms for building such classifiers. Thus in this stage, our aim is to design and build performant classifiers, through experimentation, for individual stance target-stance pairs (e.g., target-for, target-neutral, target-against) - these will be used later as part of an ensemble classifier to infer final post stances.

Due to the powerful zero-shot and few-shot abilities of auto-regressive LLMs, we used 'Starling-7B' as a base classifier model, a 7 billion parameter pre-trained auto-regressive LLM, tuned for helpfulness using reinforcement learning (Zhu et al., 2023). We experimented with others, including 'GPT-3.5-Turbo-Instruct' (OpenAI, 2023), 'Mistral-7B-Instruct-0.1' (Jiang et al., 2023a), and 'Zephyr 7B Beta' (Tunstall et al., 2023), but found that Starling-7B provided the highest accuracy, lowest cost, and its open weights access allowed fine-tuning.

As mentioned above, we seek to obtain the best possible classifiers for each combination of stance-target and stance, for use in our ensemble classifier. But we found that prompting a model for a binary of whether a text post is a specific stance or not performed worse than prompting the model for a choice out of all of the possible stances, see Appendix A.1.1 for experimental results. As such, we decided on a prompt that draws inspiration from prompts given to human annotators in previous stance detection tasks (Mohammad et al., 2016). The prompt (see Appendix A.1.2) includes a description of the stance target, which improved the accuracy of the classifier, and contextual posts, as these have both been shown to improve performance (Li et al., 2023; Zhu et al., 2022). These contextual inputs are provided alongside the stance target, and text post in question, and the model is prompted to select from the 3 stance classes as an output.

In order to further fine-tune the classifier, we needed a measure of classifier performance: we calculated the precision and recall measures for both the 'favor' and 'against' labels, and average these values, as is standard in stance detection (Mohammad et al., 2016). We used the F-beta score as our target metric (Baeza-Yates, 1999), using a beta value of 0.5, as we deemed that for this task precision is more important than recall. The reasoning for this is that if the model indicates someone has a clear opinion, we want to be sure that they do indeed have that opinion, so we need a high precision. Conversely, many expressed opinions are very subtle, and we deemed it acceptable to label as 'neutral' posts that don't clearly signal a favor or against stance.

We then improve the accuracy of our classification model with fine tuning. Initially, we experimented with chain-of-thought (CoT) methods (Zhang et al., 2023a), in-context learning (ICL) methods (Dong et al., 2022), and automated prompt-tuning methods (Li and Liang, 2021), using the DSPy library to speed up experimentation (Khattab et al., 2023). We found ICL and CoT methods were slow, and parameter-efficient fine-tuning (PEFT) methods using minimal training and validation sets resulted in the highest accuracy scores and the fastest inference speeds (Liu et al., 2022). We experimented with two variations: (1) PEFT the classifiers on all annotated stance examples, and (2) PEFT as (1), then copying that model and PEFT separate models on only the examples for a single stance target, which we dub 'two-step PEFT'. For both PEFT methods, we use 10 examples for training, and 10 for validation, with the remaining 80 from each stance target used as the test set.

The modeling work described thus far yields $N + 2$ viable classifiers, where N is the number of stance targets: the original zero-shot prompt Starling classifier, the PEFT-tuned classifier on all stances, and then two-step PEFT classifiers (one for each of the N stance targets). While one might expect that the two-step PEFT classifiers would perform best across the board, they did not. For some stance targets-label pairs, the zero-shot Starling or fine-tuned Starling classifiers performed better - which is consistent with the natural sensitivities of model training observed in other work.

To identify the best classifier for inferring the target-stance pair (t, s), we evaluated the performance of each classifier on the relevant stance target (i.e., "does this post express stance s on stance target t?") across the annotated data (on the held out test set). We refer to this classifier as $C_{(t,s)}$

and denote its cross-validation performance score as $0 \leq S_{(t,s)} \leq 1$. Note that for a given post x, $C_{(t,s)}(x)$ is either Y (yes, it does express that stance s on stance target t) or N (it does not).

We conclude this stage, then, with $3N$ classifiers: $\{C_{(t,\texttt{for})}, C_{(t,\texttt{neutral}}, C_{(t,\texttt{against})}\}$, for each of the N stance targets, t.

Final stance classification. To infer the stance of a post towards a specific stance target, we employ a simple adapted voting ensemble model: for a given post to classify and a given stance target, we run the post through each of the best classifiers for that stance target and each stance label ($\{C_{(t,\texttt{for})}, C_{(t,\texttt{neutral}}, C_{(t,\texttt{against})}\}$). We then select the final label, l_t^x, for that post by favoring the most accurate classifier that assigns its label to the post. In other words:

$B = \{s : s \in \{\texttt{for}, \texttt{against}\} \wedge C_{(t,s)} = Y\}$
if $|B| > 0$
    $l_t^x = \texttt{argmax}_{s \in B} S_{(t,s)}$
else
    $l_t^x = \texttt{neutral}$
endif

Notice that we only choose ‘neutral‘ as the label if all other labels are not assigned. In effect, we exclude neutral predictions from explicitly weighing into the label selection, as ‘neutral‘ always has the highest precision of any label (due to ‘neutral‘ being the easiest label to predict).

We then used this simple ensemble method to combine model classifications into a final stance classification for each comment. For each stance target, we classify only the posts in the respective topic.

3.3 User Stance

In this stage, we aim to obtain a measure that represents each user's stance on each stance target. However, all we currently have is a stance classification for each of their posts. Our task then, is to devise a method for aggregating the comment stance classifications into a user stance mean, and for confidence estimations, a variance. Here, we discuss that aggregation process.

Stance Aggregation. The simplest method is to simply assign -1, 0, and 1 to the stance predictions ‘against‘, ‘neutral‘, and ‘favor‘ respectively, and take the mean of these classifications as the user stance. This allows us to fairly compare the extent to which two users favor, or dis-favor, a stance. However, our classifiers not only have differing accuracy on each stance target, but they also have differing accuracy on each stance label (‘favor‘, ‘neutral‘ or ‘against‘). This means that if the classifier has higher recall at classifying text in favor of something than against it, users will seem more strongly in favor of the stance target than against it. We cannot fairly compare users favoring or dis-favoring a stance target, or compare the stance of two users on different stance targets, using this simple aggregation scheme. Even if we did not use an ensemble classifier, the differing performance of a model between labels means we need to compensate for this in any aggregation. We need to take into account the accuracy of the classifier that made the predictions.

With this in mind, we propose two methods for determining a user stance mean and variance from the predicted comment stances, using the classifier accuracy:

Weighted Mean. The simplest method is to use the weighted mean and variance of the comment stances, with weights for each comment stance being the precision of the classifier used for the comment stance prediction. We use precision as the weight, as for non-neutral predictions, this acts as a proxy for the probability that the prediction is correct. So for a set of n classification outputs x with associated classifier precision w:

$$\forall j \in \text{users}, \mu_j = \frac{\sum^n w_i x_i}{\sum^n w_i} \tag{1}$$

$$\sigma_j^2 = \left(\frac{n}{n-1}\right)\left(\frac{\sum^n w_i x_i^2}{\sum^n w_i} - \mu_w^2\right) \tag{2}$$

We can find the mean μ_j and variance σ_j^2 of a user's stance.

Through experimentation with a generative model of user stance, this method is fast, simple, and recovers the mean accurately, but gives a poor characterization of the variance, due to not factoring in the recall of the classifier. See Appendix A.2.2 for experimental results.

Inferred Mean. We therefore wanted another aggregator that could factor in the likelihood of misclassification from our classifiers. To do this we used a probabilistic generative model of the latent user stance generating the posts, which are then observed by our classifiers with error. We can then fit this model to our data, and infer the latent user stance. We will set this problem up as a maximum a posteriori probability (MAP) estimate, so that we can set a prior on our expectation of the user stance. This allows us to add an inductive bias that says: the more posts a user has, the more we'll be convinced of their stance. With a preference for fast optimization over our large dataset, we used stochastic variational inference (SVI) instead of the slower markov chain monte carlo (MCMC) inference.

We define the commenting distribution of a user as a normal distribution $\mathcal{N}(\mu, \sigma)$, where the latent continuous stance of a comment is sampled from this distribution, to represent that users produce a range of stances around their actual stance (e.g. someone favoring a target wouldn't necessarily always write content favoring their target with the same strength). Given this, the likelihood function for our latent comment stances given a user with a stance distribution is:

$$\begin{aligned} \forall j \in \text{users}, \mu_j &\sim \mathcal{N}(\mu_{loc}, \sigma_{loc}) \\ \sigma_j &\sim \log \mathcal{N}(\mu_{scale}, \sigma_{scale}) \end{aligned} \tag{3}$$

$$\forall i \in \text{posts}_j, s_i \sim \text{Normal}(\mu_j, \sigma_j) \tag{4}$$

Where $\mu_{loc}, \sigma_{loc}, \mu_{scale}, \sigma_{scale}$ are parameters to set, u_j and σ_j are the user stance variables to estimate, and s_i is the latent continuous comment stance on a given stance target. As we are measuring the posts in a discrete fashion i.e. labels of 'for', 'against', or 'neutral', we need to discretize the latent continuous comment stance:

$$q_i = \begin{cases} against & \text{if } s_i < -\frac{1}{3} \\ neutral & \text{if } -\frac{1}{3} \leq s_i \leq \frac{1}{3} \\ for & \text{if } s_i > \frac{1}{3} \end{cases} \tag{5}$$

Where q_i is the latent discretized comment stance.

At this point in the probabilistic model, we have discretized latent comment stances, but we need to fit this model to our stance predictions. So the output of our probabilistic model must be the observed comment stances, where we observe the discretized latent comment stance with the error of the classifier that observed them. We need a categorical distribution that can represent that classifier error. The closest thing we have to the true categorical distribution of the classifier error given a true comment stance, is the column of the confusion matrix for the true comment stance, obtained at test time. We can normalize this column to get an approximation of the classifier's categorical error distribution, as has been used in similar methods previously (Kerrigan et al., 2021).

$$\mathrm{x}_i \sim \text{Categorical}(P(X|Q = q_i)) \tag{6}$$

Where x_i is the observed comment stance.

With that, our probabilistic model of the likelihood function of our data generation process is complete. To optimize the model, we need to approximate the posterior of this process, with the variational distribution. To produce a variational distribution, we use the confusion matrix of the model to approximate the $P(Q|X = x_i)$ probability: the probability of the true latent comment stance, given the comment stance observations. We model the latent variables μ_j and σ_j in the variational distribution as $Delta$ distributions, for MAP inference. Given this likelihood function, and the variational distribution, we can find the variables μ_j and σ_j which most likely gave us our data by maximising the evidence lower bound (ELBO). See Appendix A.2.1 for training details.

With this method in hand, we can aggregate the comment stance predictions into our final user stance mean and variance, μ_j and σ_j. We evaluated this method using synthetic data, and found that while it apportioned probability mass more accurately for users with fewer data points and classifiers with error, this came at a cost of the mean of the inferred normal systematically underestimating the true user stance, due to the distributed probability mass. This makes the method suitable for downstream applications that can use this information well (i.e. probabilistic models), but less appropriate for applications which can only factor in the mean. See Appendix A.2.2 for experimental results and discussion of this evaluation.

4 Experiments

We validated the proposed method by looking at Canadian political opinion dynamics on Reddit. After detecting stances over the Reddit corpus collected using the proposed method, we evaluated the extent to which inferred user stances reproduce known opinion polling results and temporal opinion trends.

4.1 Data, stance-targets, and inference

Data. Using the Pushshift dataset, we collected all 2022 content from the 4 largest Canadian-centric subreddits: 'r/canada', 'r/vancouver', 'r/ontario', and 'r/toronto' (Baumgartner et al., 2020).

Stance Target	Description	Fleiss' Kappa
Vaccine Mandates	Laws requiring personal use of COVID-19 vaccines	0.53
Renter Protections	Laws protecting the rights of people renting housing	0.23
NDP	The NDP Party of Canada	0.59
Liberals	The Liberal Party of Canada	0.56
Conservatives	The Conservative Party of Canada	0.63
Gun Control	Laws regulating the use of firearms	0.55
Drug Decriminalization	Policy decriminalizing illegal drugs	0.46
Liberal Immigration Policy	Laws favoring more immigration	0.59
Canadian Aid to Ukraine	Government financial and military aid to Ukraine	0.623
French Language Laws	Laws mandating the use of the French language	0.547

Table 1: Our chosen stance targets, descriptions for each stance target for reader context, and the Fleiss' Kappa statistic for the interannotator agreement of the annotations from our annotators (Gwet, 2008).

	Prec	Rec	F1
1. Zero Shot (Avg)	0.42	0.61	0.45
2. Two-step PEFT (Avg)	0.60	0.51	0.52
3. PEFT (Avg)	0.59	0.52	0.53
4. Ensemble (Avg)	0.74	0.64	0.65
5. Always Favor	0.06	0.50	0.10
6. Annotator	0.74	0.84	0.77
7. Twitter	0.93	0.91	0.92
8. Reddit	-	-	0.594

Table 2: Mean macro precision, recall, and F1 (excluding neutral label) of techniques used, including methods composing the ensemble (1-3), the ensemble (4), baselines (including human performance) (5-6), and prior work (7-8). Prior work is selected from previous comparable work, and included to contextualize the performance of our classifiers with contemporaneous work, but note these numbers are not for the same dataset. We report the F1 metric, as this is the one reported by prior work. Prior work: (Samih and Darwish, 2021) (Twitter), and (Kim et al., 2023) (Reddit)

Stance target selection. For topic modeling, we used a higher than recommended number of neighbours parameter for UMAP of 30, and a minimum cluster size of 0.1% of the dataset for HDBSCAN, to find larger topics that covered coarse-grained political issues. This left us with 80 clusters aligned with political issues, where 51% of texts were considered outliers from these clusters. The actual selection of the stance targets was, frankly, subjective - though we expect this would be the case in empirical opinion studies as well. In the end, we selected the stance targets shown in Table 1, where we also include a description of the stance target. The posts in the topics associated with these chosen stance targets represented 16% of the text in our dataset.

Annotation and modeling. We took 100 samples from topics linked to our selected stance targets and double-annotated them using the annotation procedure described in Section 3.2. We ran the Fleiss' Kappa interannotator agreement statistic on the annotations, and present them in Table 1.

We tried training our classifiers with a number of different methods, and we report the metrics from each of those methods in Table 2. Our two methods of parameter efficient fine-tuning worked out to produce the classifiers with the highest accuracy. We then used the highest accuracy individual classifiers for our ensemble method.

Stance detection. We ran the classifier on all the comments from each topic deemed close to our targets for all users who had at least 5 comments (as a rough proxy for ensuring we had enough classifications to obtain a reasonable stance signal). We then ran our user stance aggregation methods, to obtain user stance scores for each user on each stance target.

4.2 Investigations

There are many political behaviour investigations we can perform with the data available from this work, both to learn about views on political issues in Canadian politics, and validate the accuracy of our results. We therefore ask 3 initial research questions (RQs):

1. **RQ1: Can we find the political issue opinion correlation between sets of issues, and which are the strongest among them?** We

start by taking a static, big picture view of online opinions in Canadian politics.

2. **RQ2: What can we learn about polarization in Canadian political discussion from this data?** If there are correlations in political issue opinions, this indicates there is polarization (political sorting). Can we quantify that polarization, and the polarization of subsets of the data?
3. **RQ3: Do we observe fluctuations in political issue opinion over time, and do these line up with potentially explanatory events?** Political issues change over time, and with them, the opinions of the people. If we can observe this with confidence, it will provide a strong tool for viewing and understanding reactions to public policy.

4.2.1 RQ1: Static user opinions trends

It is well-established that opinions on certain topics are correlated (Baldassarri and Gelman, 2008). This is the question we prepare to investigate here. In Figure 1, we see selected scatter plots of users' opinions on multiple stance targets (See Fig. 4 for all scatter plots and distributions). These trends in these scatter plots can be inspected to assess the correlations that may exist between different stance targets. We used the inferred mean method for the user stance data in this experiment.

We fit a weighted least squares models to each of the opinion comparisons (Seber and Lee, 2012), and include the resulting correlation, p-value, and R^2 value in Figure 1. Note many people have only posted opinionated ('favor' or 'against', not 'neutral') comments on one stance target or the other, and therefore we get a prominent 'cross' on the scatter, combined with an obvious cluster of users who are opinionated on both stances. In order to see the correlation of this opinionated cluster of users, we re-calculate the correlation without users who have not posted an opinionated comment. We report these correlations with the highest R^2 values in Table 3, where we see numerous strong correlations between stance targets.

Inferred stances reveal strong correlations in user stance between targets. Crucially, inter-target stance correlations line up with Canadian political party platforms - that is, users who support a political party are likely to aligned with the party's political platform (CBC, 2021). For example, we find strong opposite correlations between Liberals and Gun Control, and Conservatives and Gun Control.

Target A	Target B	ρ	R^2
Vacc. Mandates	Gun Control.	0.29	0.12
Vacc. Mandates	Immigr. Pol.	0.16	0.13
Liberals	Gun Control	0.29	0.15
Liberals	Immigr. Pol.	0.20	0.15
Gun Control	Immigr. Pol.	0.28	0.19

Table 3: User stance correlations issue pairs, including only correlations with $R^2 >= 0.1$ for brevity.

For many political issue pairings, users are polarized - that is, they are self-sorting into ideological camps (Baldassarri and Gelman, 2008). Curiously, some of the highest correlations and most predictive relationships between political issue stances were non-party stance targets, suggesting political issues are more polarizing than party allegiances.

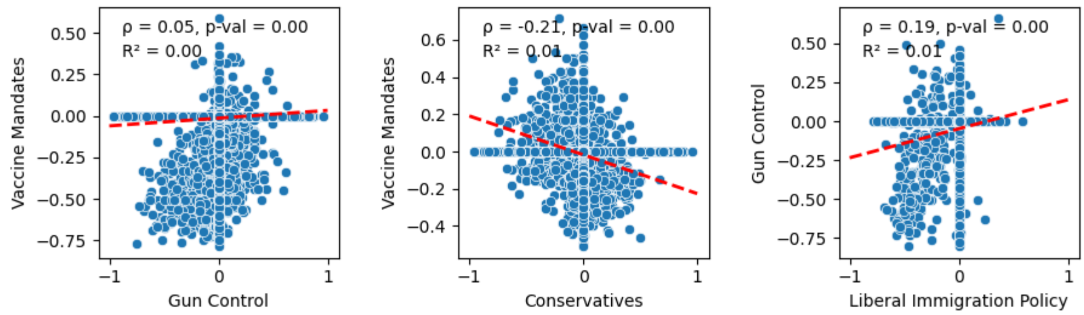
4.2.2 RQ2: Polarization.

We want to dive more into the topic of polarization, due to the evidence we see for it in the correlation data, and it's consequences for society (Klein, 2020). Using this static opinion data, we compute polarization measures for each subset of stance targets (Gubanov et al., 2021). Imagine a room of opinionated people - if they all know each others full spectrum of opinions, symmetric polarization gives the extent to which two separate groups of disagreeing people would form. However, if we were to only let them know a subset of each others opinions, then asymmetric polarization tells us which subsets of opinions produce the most and least divided rooms.

Using asymmetric polarization, we find that gun control and liberal immigration policy are among the most polarized stance targets, and that renter protections and Canadian aid to Ukraine are the least polarized stance targets. The measure tells us that 'r/canada' is the most polarized subreddit with a symmetric polarization measure of 0.11, and 'r/vancouver' is the least polarized subreddit with a measure of 0.04, indicating that national political discussion is more polarized than provincial political discussion.

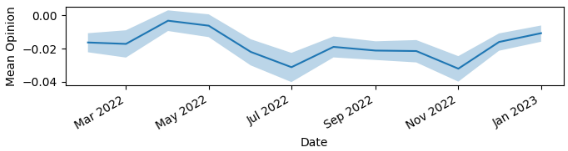
4.2.3 RQ3: Temporal trends

Our method provides annotations for timestamped social media data. As a result, we also obtain information about how user opinions change, en mass, over time. To the best of our knowledge, this level of population opinion polling is unmatched for political issues beyond party opinion and major topi-

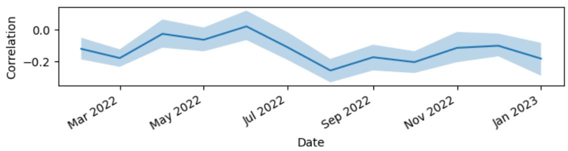


(a) Vaccine mandates and gun control user stance scatter plot.
(b) Vaccine mandates and conservatives user stance scatter plot.
(c) Gun control and liberal immigration policy user stance scatter plot.

Figure 1: Scatter plots compare user stances on two different stance targets. We fit a weighted least squares method to the data (using the inverse of the sum of the user stance variances as a weight), to find the correlation of the data.



(a) Vaccine mandates opinion timeline.



(b) Vaccine mandates and conservatives opinion correlation timeline.

Figure 2: Movement of stance opinion and correlation over 2022. The shaded bars indicate the confidence intervals of the moving statistic, as determined by the bootstrap method (Efron and Tibshirani, 1994).

cal issues (e.g. vaccine mandates).

To do this, we look at how the aggregated user opinions shift over each month of 2022. We constructed figures of the change in mean user opinion and correlation of user opinions on a month by month basis, complete with confidence intervals found via the bootstrap method (Efron, 1992). We include a sample of these in Figure 2.

These temporal opinion shifts reflect known inflection points in policy and public opinion. For example, looking at vaccine mandates opinion trends (Fig. 2a), the first data point we have is in April 2022, when the overall stance on vaccine mandates reaches a high, coinciding with polling suggesting that the public was less worried about COVID-19 (Coletto and Anderson, 2022). By July, the opinion for vaccine mandates drops as reports circulate about vaccine skepticism (Institute, 2022). The correlation between conservatives stance and vaccine mandates stance also drops, indicating conservatives become more likely to be against vaccine mandates (or vice versa), (Fig. 2b) following reports about vaccine hesitancy, and talk of ending vaccine mandates (Lavery, 2022; Boutilier, 2022).

5 Discussion

In this work, we aimed to provide a complete template method for moving from raw social media data and inferring user stances across multiple stance targets. While the method itself primarily assembles existing methods, it does so in systematic a way that, to our knowledge, have not been attempted before. Moreover, we have also contributed a novel approach to rendering a user's stance on multiple stance targets comparable. Applied to Reddit data, the user stance trends our method yields reflect known and notable behaviour.

Future work These findings collectively point to the utility of the method we proposed in this paper. But there is a great deal of work to be done to improve these methods. For the first part of our methodology, automated techniques that can select the most salient stance targets would reduce any potential bias in manually choosing them, and initial methods for this have started to appear (Paschalides et al., 2021). Our methodology could also be extended to non-text based platforms by means of large multimodal models (LMMs) (Liu et al., 2023). And finally, there is much analysis work possible with the user stance signals, where we could use more sophisticated modelling techniques to uncover more complex opinion dynamics.

6 Ethical Statement

The method described in this work has a strong privacy violating potential. Although all of the text used in this work is publicly available for anyone with an internet connection, rapid progression of derived features from social trace data makes informed consent impossible. Just as this method allows us to understand large scale human behaviour, it also makes it possible to track and predict individual user's stances. However, we believe these methods have strong democratic potential for better understanding population perspectives. Therefore we believe it is critical to always focus our analysis on this data in the aggregate, and use this data to understand large scale trends as opposed to investigating specific users.

7 Limitations

We would highlight that our template does not capture the only assembly of existing methods to achieve its aim. Our objective here is to provide *a* credible and reproducible way of measuring multi-target user stance, and we invite future work to improve and contribute other frameworks for this task.

A drawback of our method is the disconnection between topics and opinion dimensions. The first disadvantage of this is that choosing the stance target we examine in the topic of interest is done manually, and this opens this choice up to being either an opinion dimension which is either not heavily discussed and therefore not representative of the analysed discussion, or a misleading dimension of disagreement that is better associated with a larger, more delineated discussion. Future work should use an automated method to discover these stance targets (Paschalides et al., 2021).

Another, related, problem in this method is how we can understand the idea of a 'neutral' stance in each stance category. For some topics, many comments associated with that topic are discussing the stance target in question, such as comments associated with a gun topic discussing gun control. However, for other topics, like Canadian political parties, a smaller share of the comments in this topic are discussing the new democratic party (NDP), so the NDP get a disproportionately higher number of 'neutral' comments.

Finally there is more work to be done on the user stance aggregation process, namely to further validate the likelihood function and variational distribution that we used. We experimented with using a beta distribution to represent skew in the potential user stance, but had difficulties with limiting the distribution parameters to realistic user stance distributions, so more work can be done here.

References

TM Abeysinghe. 2023. Utilizing multi-modal weak signals to improve user stance inference in social media.

Emily Allaway and Kathleen McKeown. 2023. Zero-shot stance detection: Paradigms and challenges. *Frontiers in Artificial Intelligence*, 5:1070429.

Ali Almadan, Mary Lou Maher, and Jason Windett. 2023. Stance detection for gauging public opinion: A statistical analysis of the difference between tweet-based and user-based stance in twitter. In *Future of Information and Communication Conference*, pages 358–374. Springer.

R Baeza-Yates. 1999. Modern information retrieval. *Addison Wesley google schola*, 2:127–136.

Delia Baldassarri and Andrew Gelman. 2008. Partisans without constraint: Political polarization and trends in american public opinion. *American Journal of Sociology*, 114(2):408–446.

Jason Baumgartner, Savvas Zannettou, Brian Keegan, Megan Squire, and Jeremy Blackburn. 2020. The pushshift reddit dataset. In *Proceedings of the international AAAI conference on web and social media*, volume 14, pages 830–839.

David M Blei, Andrew Y Ng, and Michael I Jordan. 2003. Latent dirichlet allocation. *Journal of machine Learning research*, 3(Jan):993–1022.

Ana-Maria Bliuc, Craig McGarty, Katherine Reynolds, and Daniela Muntele. 2007. Opinion-based group membership as a predictor of commitment to political action. *European journal of social psychology*, 37(1):19–32.

Alex Boutilier. 2022. 'this will be pierre poilievre's party': Conservatives reckon with a new direction. *Global News*.

CBC. 2021. How do the main parties compare on these issues? https://newsinteractives.cbc.ca/elections/federal/2021/party-platforms/. Accessed: 2024-05-17.

David Coletto and Bruce Anderson. 2022. Covid anxiety moderate to low: 6th wave hasn't caused a spike in fear. *Abacus Data*.

Iain J Cruickshank and Lynnette Hui Xian Ng. 2023. Use of large language models for stance classification. *arXiv preprint arXiv:2309.13734*.

Kareem Darwish, Peter Stefanov, Michaël Aupetit, and Preslav Nakov. 2020. Unsupervised user stance detection on twitter. In *Proceedings of the International AAAI Conference on Web and Social Media*, volume 14, pages 141–152.

Qingxiu Dong, Lei Li, Damai Dai, Ce Zheng, Zhiyong Wu, Baobao Chang, Xu Sun, Jingjing Xu, and Zhifang Sui. 2022. A survey for in-context learning. *arXiv preprint arXiv:2301.00234*.

Bradley Efron. 1992. Bootstrap methods: another look at the jackknife. In *Breakthroughs in statistics: Methodology and distribution*, pages 569–593. Springer.

Bradley Efron and Robert J Tibshirani. 1994. *An introduction to the bootstrap*. Chapman and Hall/CRC.

Maarten Grootendorst. 2022. Bertopic: Neural topic modeling with a class-based tf-idf procedure. *arXiv preprint arXiv:2203.05794*.

Dmitry A Gubanov, Ilya V Petrov, and Alexander G Chkhartishvili. 2021. Multidimensional model of opinion dynamics in social networks: polarization indices. *Automation and Remote Control*, 82:1802–1811.

Kilem Li Gwet. 2008. Computing inter-rater reliability and its variance in the presence of high agreement. *British Journal of Mathematical and Statistical Psychology*, 61(1):29–48.

Angus Reid Institute. 2022. Covid-19: Half want boosters asap, but two-in-five among vaccinated say they're not sold on another shot.

Joshua Introne. 2023. Measuring belief dynamics on twitter. In *Proceedings of the International AAAI Conference on Web and Social Media*, volume 17, pages 387–398.

Albert Q Jiang, Alexandre Sablayrolles, Arthur Mensch, Chris Bamford, Devendra Singh Chaplot, Diego de las Casas, Florian Bressand, Gianna Lengyel, Guillaume Lample, Lucile Saulnier, et al. 2023a. Mistral 7b. *arXiv preprint arXiv:2310.06825*.

Julie Jiang, Xiang Ren, and Emilio Ferrara. 2023b. Retweet-bert: political leaning detection using language features and information diffusion on social networks. In *Proceedings of the International AAAI Conference on Web and Social Media*, volume 17, pages 459–469.

Gavin Kerrigan, Padhraic Smyth, and Mark Steyvers. 2021. Combining human predictions with model probabilities via confusion matrices and calibration. *Advances in Neural Information Processing Systems*, 34:4421–4434.

Omar Khattab, Arnav Singhvi, Paridhi Maheshwari, Zhiyuan Zhang, Keshav Santhanam, Sri Vardhamanan, Saiful Haq, Ashutosh Sharma, Thomas T Joshi, Hanna Moazam, et al. 2023. Dspy: Compiling declarative language model calls into self-improving pipelines. *arXiv preprint arXiv:2310.03714*.

Rachel Kim, Veniamin Veselovsky, and Ashton Anderson. 2023. Capturing dynamics in online public discourse: A case study of universal basic income discussions on reddit. *arXiv preprint arXiv:2312.09611*.

Diederik P Kingma and Jimmy Ba. 2014. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*.

Ezra Klein. 2020. *Why we're polarized*. Simon and Schuster.

Irelyne Lavery. 2022. Poilievre popular among conservative voters, but not all canadians feel the same: poll. *Global News*.

Xiang Lisa Li and Percy Liang. 2021. Prefix-tuning: Optimizing continuous prompts for generation. *arXiv preprint arXiv:2101.00190*.

Yingjie Li, Chenye Zhao, and Cornelia Caragea. 2021. Improving stance detection with multi-dataset learning and knowledge distillation. In *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing*, pages 6332–6345.

Yupeng Li, Haorui He, Shaonan Wang, Francis CM Lau, and Yunya Song. 2023. Improved target-specific stance detection on social media platforms by delving into conversation threads. *IEEE Transactions on Computational Social Systems*.

Haokun Liu, Derek Tam, Mohammed Muqeeth, Jay Mohta, Tenghao Huang, Mohit Bansal, and Colin A Raffel. 2022. Few-shot parameter-efficient fine-tuning is better and cheaper than in-context learning. *Advances in Neural Information Processing Systems*, 35:1950–1965.

Haotian Liu, Chunyuan Li, Yuheng Li, and Yong Jae Lee. 2023. Improved baselines with visual instruction tuning. *arXiv preprint arXiv:2310.03744*.

Leland McInnes and John Healy. 2017. Accelerated hierarchical density based clustering. In *2017 IEEE international conference on data mining workshops (ICDMW)*, pages 33–42. IEEE.

Leland McInnes, John Healy, and James Melville. 2018. Umap: Uniform manifold approximation and projection for dimension reduction. *arXiv preprint arXiv:1802.03426*.

Saif Mohammad, Svetlana Kiritchenko, Parinaz Sobhani, Xiaodan Zhu, and Colin Cherry. 2016. Semeval-2016 task 6: Detecting stance in tweets. In *Proceedings of the 10th international workshop on semantic evaluation (SemEval-2016)*, pages 31–41.

OpenAI. 2023. Gpt-3.5 turbo. https://platform.openai.com/docs/models/gpt-3-5-turbo. Accessed: 2024-05-16.

Demetris Paschalides, George Pallis, and Marios D Dikaiakos. 2021. Polar: a holistic framework for the modelling of polarization and identification of polarizing topics in news media. In *Proceedings of the 2021 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining*, pages 348–355.

Ammar Rashed, Mucahid Kutlu, Kareem Darwish, Tamer Elsayed, and Cansın Bayrak. 2021. Embeddings-based clustering for target specific stances: The case of a polarized turkey. In *Proceedings of the International AAAI Conference on web and social media*, volume 15, pages 537–548.

Nils Reimers and Iryna Gurevych. 2019. Sentence-bert: Sentence embeddings using siamese bert-networks. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing*. Association for Computational Linguistics.

Maud Reveilhac, Stephanie Steinmetz, and Davide Morselli. 2022. A systematic literature review of how and whether social media data can complement traditional survey data to study public opinion. *Multimedia tools and applications*, 81(7):10107–10142.

Younes Samih and Kareem Darwish. 2021. A few topical tweets are enough for effective user stance detection. In *Proceedings of the 16th Conference of the European Chapter of the Association for Computational Linguistics: Main Volume*, pages 2637–2646.

George AF Seber and Alan J Lee. 2012. *Linear regression analysis*. John Wiley & Sons.

Lewis Tunstall, Edward Beeching, Nathan Lambert, Nazneen Rajani, Kashif Rasul, Younes Belkada, Shengyi Huang, Leandro von Werra, Clémentine Fourrier, Nathan Habib, et al. 2023. Zephyr: Direct distillation of lm alignment. *arXiv preprint arXiv:2310.16944*.

Jenq-Haur Wang and Kuan-Ting Chen. 2021. Aggregating user-centric and post-centric sentiments from social media for topical stance prediction. In *Proceedings of the 33rd Conference on Computational Linguistics and Speech Processing (ROCLING 2021)*, pages 226–235.

Bowen Zhang, Xianghua Fu, Daijun Ding, Hu Huang, Yangyang Li, and Liwen Jing. 2023a. Investigating chain-of-thought with chatgpt for stance detection on social media. *arXiv preprint arXiv:2304.03087*.

Hong Zhang, Haewoon Kwak, Wei Gao, and Jisun An. 2023b. Wearing masks implies refuting trump?: Towards target-specific user stance prediction across events in covid-19 and us election 2020. In *Proceedings of the 15th ACM Web Science Conference 2023*, pages 23–32.

Zijian Zhou, Shuoyu Hu, Kai Yang, and Haizhou Wang. 2023. Tksp: Long-term stance prediction for social media users by fusing time series features and event dynamic evolution knowledge. In *International Conference on Knowledge Science, Engineering and Management*, pages 406–421. Springer.

Banghua Zhu, Evan Frick, Tianhao Wu, Hanlin Zhu, and Jiantao Jiao. 2023. Starling-7b: Improving llm helpfulness harmlessness with rlaif.

Lixing Zhu, Yulan He, and Deyu Zhou. 2020. Neural opinion dynamics model for the prediction of user-level stance dynamics. *Information Processing & Management*, 57(2):102031.

Qinglin Zhu, Bin Liang, Jingyi Sun, Jiachen Du, Lanjun Zhou, and Ruifeng Xu. 2022. Enhancing zero-shot stance detection via targeted background knowledge. In *Proceedings of the 45th international ACM SIGIR conference on research and development in information retrieval*, pages 2070–2075.

A Appendix

A.1 Prompt

A.1.1 Prompt Type

We experimented with two main prompt types:

- **All stances**: One question: is the stance favor, against, or neutral?
- **Ask each stance**: One prompt asking if the post favors the stance target, one prompt asking if the post is against the stance target.

See Table 4 for the best performing result from each prompt type. We see that the single question outperforms the 'ask for each stance' method.

	Prec	Rec	F1
All stances	0.42	0.61	0.45
Ask each stance	0.40	0.31	0.29

Table 4: Best performing results from each prompt type. See

A.1.2 Final Prompt

We used the following prompt for our stance classifier, that we arrived at after manual experimentation. Any curly brackets demarked variables not preceded by a $ are templated for the parameter used in that example:

> Predict the stance of the comment towards {target_opinion}. Here is an explanation of what we mean by {target_opinion}: {target_explanation} If the comment is directly or indirectly in favor of {target_opinion}, or opposing

or critizing something opposed to {target_opinion}, then the stance should be favor. If the comment is directly or indirectly against {target_opinion}, or opposing or critizing something in favor of {target_opinion}, then the stance should be against. If the comment is discussing something irrelevant to {target_opinion}, or if it is unclear what the stance is, then the stance should be neutral.

Post: The post being commented on, may be useful in determining what the comment is discussing.
Parent Comment: The parent comment being replied to, may be useful in determining the context of the comment.
Comment: The comment to determine the opinion of.
Stance: The stance of the comment is ${favor, neutral, or against}

Post: {post}
Parent Comment: {parent_comment}
Comment: {comment}
Stance: The stance of the comment is

After additional experimentation, we used different wordings for the stance targets listed in the paper above to improve accuracy, and wrote out stance target descriptions. These alternate wordings and descriptions can be seen in the project GitHub repository github.com/bendavidsteel/user-stance-discovery. We used the same stance target descriptions for our annotators.

A.2 User Stance Estimation

A.2.1 Training

We used the Pyro library [1] to build and train the probabilistic model, using the clipped Adam optimizer (Kingma and Ba, 2014) for 1000 epochs, with an initial learning rate of 0.1 decaying to 0.001 over the training run.

A.2.2 Experiments

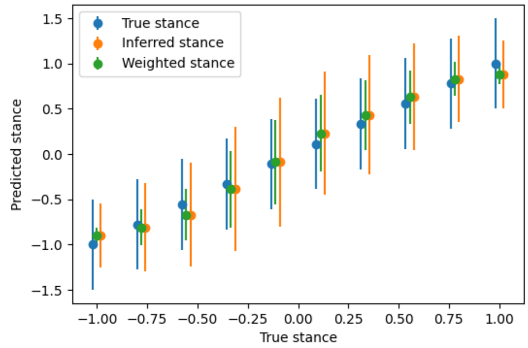
We used a simple generative model of a user to generate synthetic data to test the user stance inference with. The model has a latent user stance μ (represented as a scalar between -1 and 1), and user stance variance σ^2 (the variance of of comment stances, to model consistency of stance), to parameterize a user stance distribution. We can then draw N comments from this normal distribution, model classifying them with error via a categorical distribution defined by a specific precision and recall (to simulate a classifier with error), to produce the final synthetic data. We show the results from the experiments we did in Figure 3.

We can see that for this generative model, though the weighted mean method recovers the true generative model user stance mean most accurately, it fails to apportion probability mass well when there are fewer data points, or the precision/recall is lower, resulting in inaccurate variances. The inferred mean method can more accurately apportion probability mass, but systematically underestimates the true user stance for classifiers with lower recall and precision. This is because probability mass is placed more around 0, the centre of the user stance domain, to account for possible neutral posts, and there is no probability mass placed beyond 1 or -1, as that is the limit of our discrete comment stances. The normal distribution therefore correspondingly moves towards 0. This could be improved by modelling the user stance as a beta distribution, as the skew can accommodate for the possible neutral posts. However, in tests, we found it difficult to constrain the beta distribution to reasonable user stance distributions. More work is necessary here.

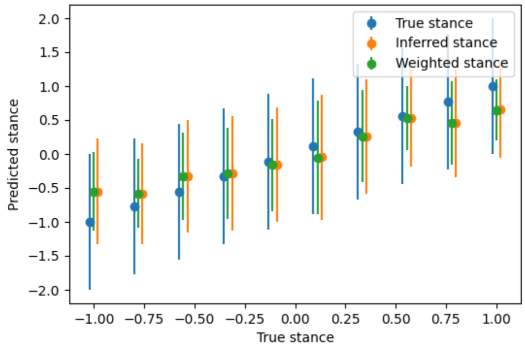
A.3 Static User Opinions

We include in Fig. 4 a general overview of static user stances on the stance targets we cover in this work.

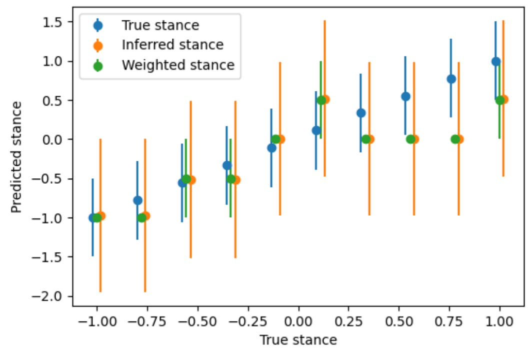
[1] https://github.com/pyro-ppl/pyro



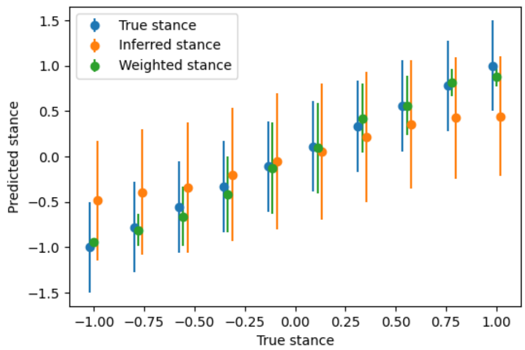
(a) Experiment with $N = 100$, $\sigma^2 = 0.5$, $Precision = 1.0$, and $Recall = 1.0$.



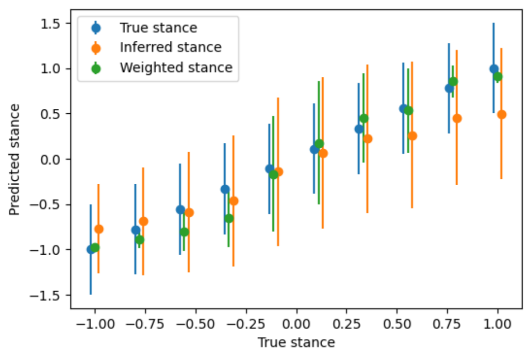
(b) Experiment with $N = 100$, $\sigma^2 = 1.0$, $Precision = 1.0$, and $Recall = 1.0$.



(c) Experiment with $N = 2$, $\sigma^2 = 0.5$, $Precision = 1.0$, and $Recall = 1.0$. Note that the weighted user stance method reports a variance of 0 due to the small number of data points all being in the same class, whereas the inference method reports a large variance, indicating there's still great uncertainty in the user stance.



(d) Experiment with $N = 100$, $\sigma^2 = 0.5$, $Precision = 0.559$, and $Recall = 0.559$. Note that the inference method tends to systematically underestimate the true user stance, due to the high classifier error ensuring user stance probability mass is spread out.



(e) Experiment with $N = 100$, $\sigma^2 = 0.5$, $Precision = 0.718$, and $Recall = 0.396$.

Figure 3: Results from validation of our weighted and inferred user stance methods in Section 3.3. For all experiments, we generate N comments from 10 synthetic users with μ ranging from -1 to 1, and plot the predicted user stance against the true user stance. We vary the σ^2, N, precision and recall by experiment to test the methods in different situations.

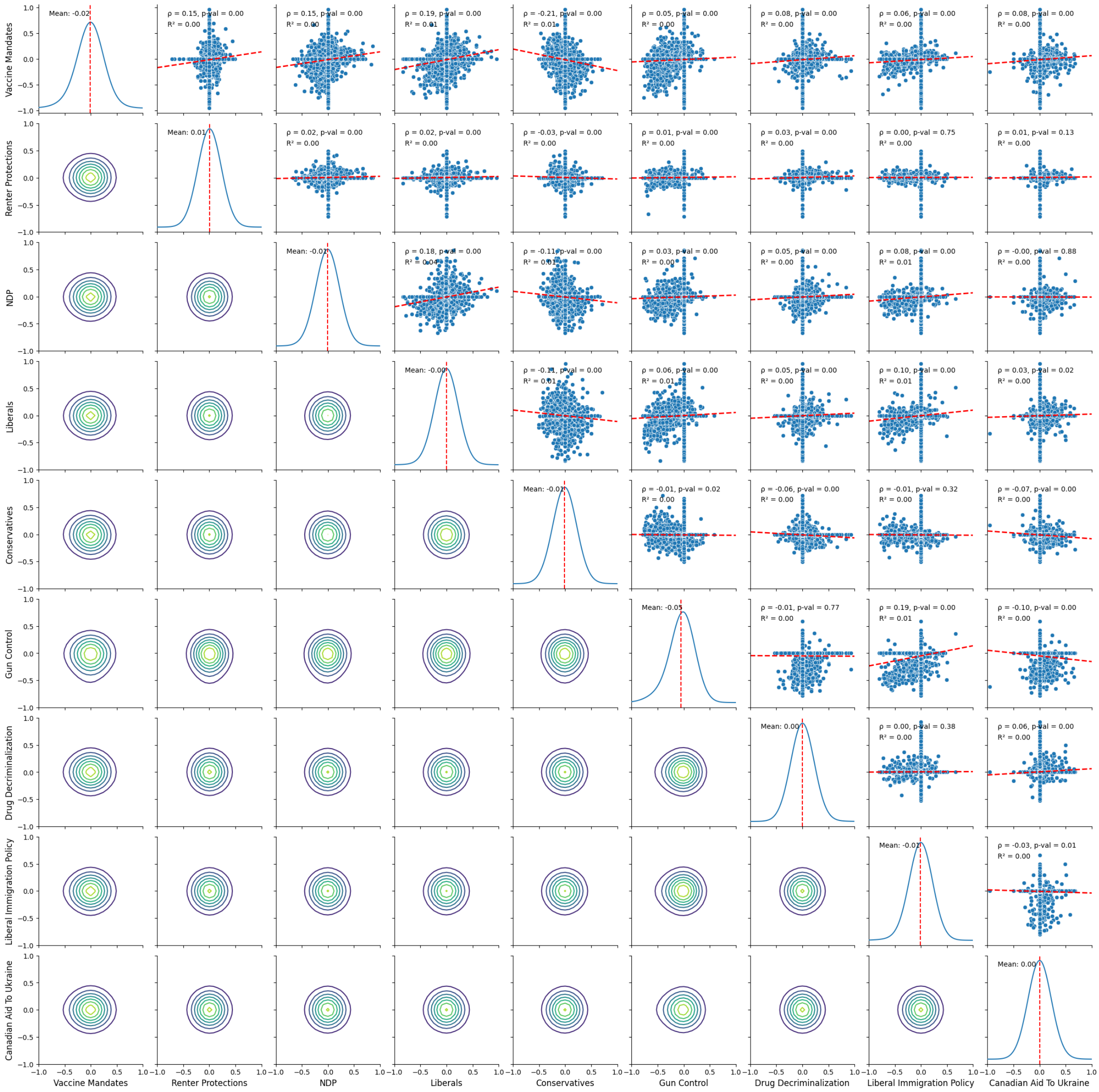


Figure 4: Plot exploring user stances for each stance target. The diagonal plot shows the inferred distribution of each stance target, determined by summing all of the normal distributions we determine through SVI. We include the mean of this distribution, as a dashed line. The below diagonal plots show the inferred distribution of paired stance dimensions, showing where users are likely to fit within a bi-dimensional stance space, determined by summing all inferred bivariate user stance normal distributions. The above diagonal plots show a scatter graph of user stances in each bivariate space, including a correlation determined by the weighted least squares method.