Investigating Human Values in Online Communities

Nadav Borenstein, Arnav Arora, Lucie-Aimée Kaffee, Isabelle Augenstein

Department of Computer Science
University of Copenhagen
Denmark

nb@di.ku.dk | aar@di.ku.dk | lucie-aimee.kaffee@hpi.de | augenstein@di.ku.dk

ArXiv Link

UNIVERSITY OF COPENHAGEN

Motivation

- Studying human values is key to understanding the behaviours and preferences of communities.
- However, traditional methods are slow and expensive, mostly involving self-reporting questionaries and manually annotating data.
- This work proposes a computational method to study Schwartz values on Reddit and analyse online communities at scale.

Value	Description
Power	Social status and prestige, control or dominance over people and resources
Achievement	Personal success through demonstrating competence according to social standards.
Hedonism	Pleasure and sensuous gratification for oneself.
Stimulation	Excitement, novelty, and challenge in life.
Self-direction	Independent thought and action-choosing, creating, exploring.
Universalism	Understanding, appreciation, tolerance, and protection for the welfare of all people and for nature.
Benevolence	Preservation and enhancement of the welfare of people with whom one is in frequent personal contact.
Tradition	Respect, commitment, and acceptance of the customs and ideas that traditional culture or religion provide.
Conformity	Restraint of actions, inclinations, and impulses likely to upset or harm others and violate social expectations or norms.
Security	Safety, harmony, and stability of society, of relationships, and of self.

Val	Subreddits
AC	r/startups, r/resumes, r/xboxachievements
BE	r/Adoption, r/BPDlovedones, r/Petloss
CO	r/policebrutality, r/HOA, r/BadNeighbors,
HE	r/FreeCompliments, r/transpositive, r/cozy
PO	r/debtfree, r/geopolitics, r/dividends
SE	r/GunsAreCool, r/worldevents, r/CombatFootage
SD	r/antidepressants r/DebateReligion, r/TrueUnpopularOpinion, r/nutrition
ST	r/crossdressing, r/Hobbies, r/NailArt
TR	r/religion, r/AskAPriest, r/atheism
UN	r/AskFeminists, r/IsraelPalestine, r/climatechange

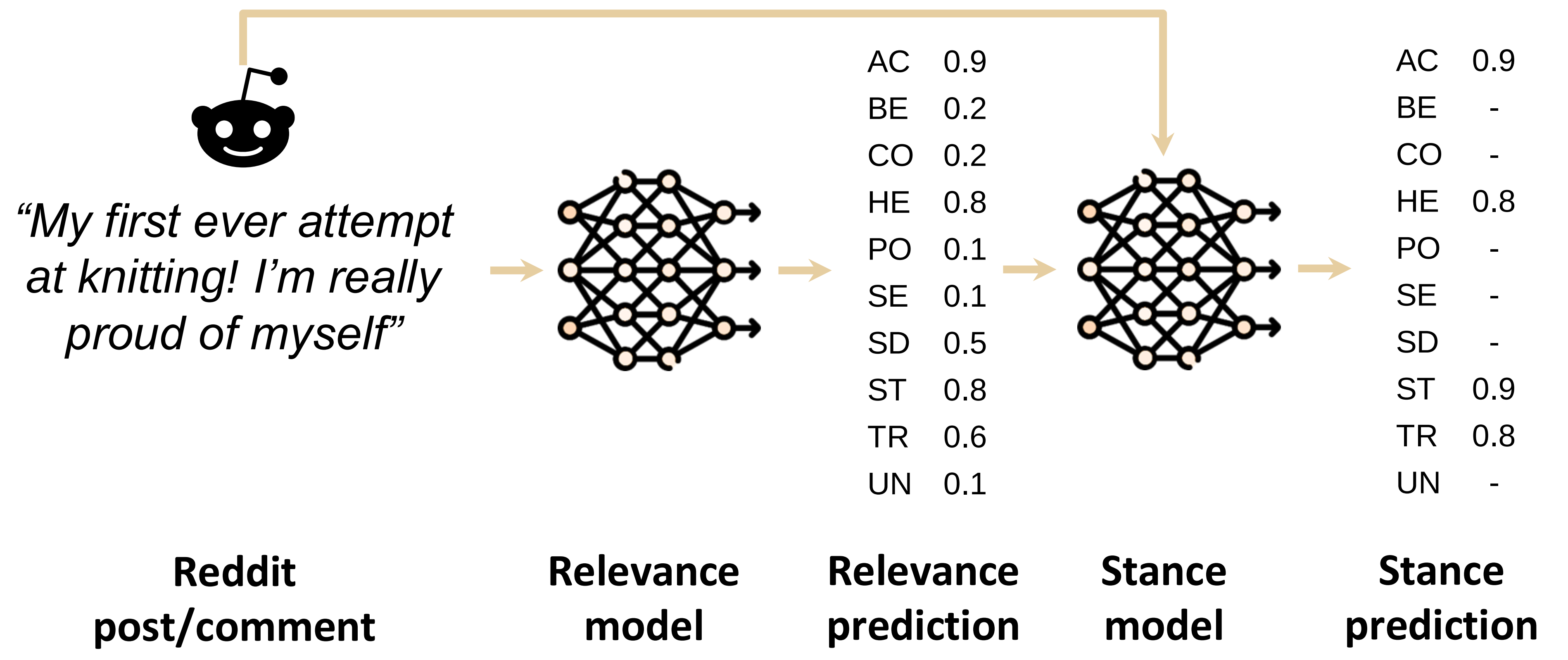
Subreddits with the highest expression of each of the ten Schwartz values. The stance of **Green** subreddits towards the value is positive (above 0.2), whereas **Red** indicates a negative stance (below 0.2). **Blue** represents neutral.

Method

Datasets and training

- **Reddit data**: 9M posts and comments from the top 10k subreddits.
- **Training data:** ValueNet and ValueArg. Manually annotated datasets of single-sentence statements.
- **Training:** Fine-tune two DeBERTa models to extract value relevance and stance.

Inference



"My first ever attempt at knitting! I'm really proud of myself"

	Relevance prediction	Stance prediction
AC	0.9	0.9
BE	0.2	-
CO	0.2	-
HE	0.8	0.8
PO	0.1	-
SE	0.1	-
SD	0.5	-
ST	0.8	0.9
TR	0.6	0.8
UN	0.1	-

Reddit post/comment → Relevance model → Relevance prediction → Stance model → Stance prediction

Assigning values to entire subreddit by averaging posts and comments:

$$u_{\text{rel}}(\mathcal{S}) = \frac{1}{|\mathcal{S}|}\sum_{i\in\mathcal{S}} u_{\text{rel}}(c_i)$$

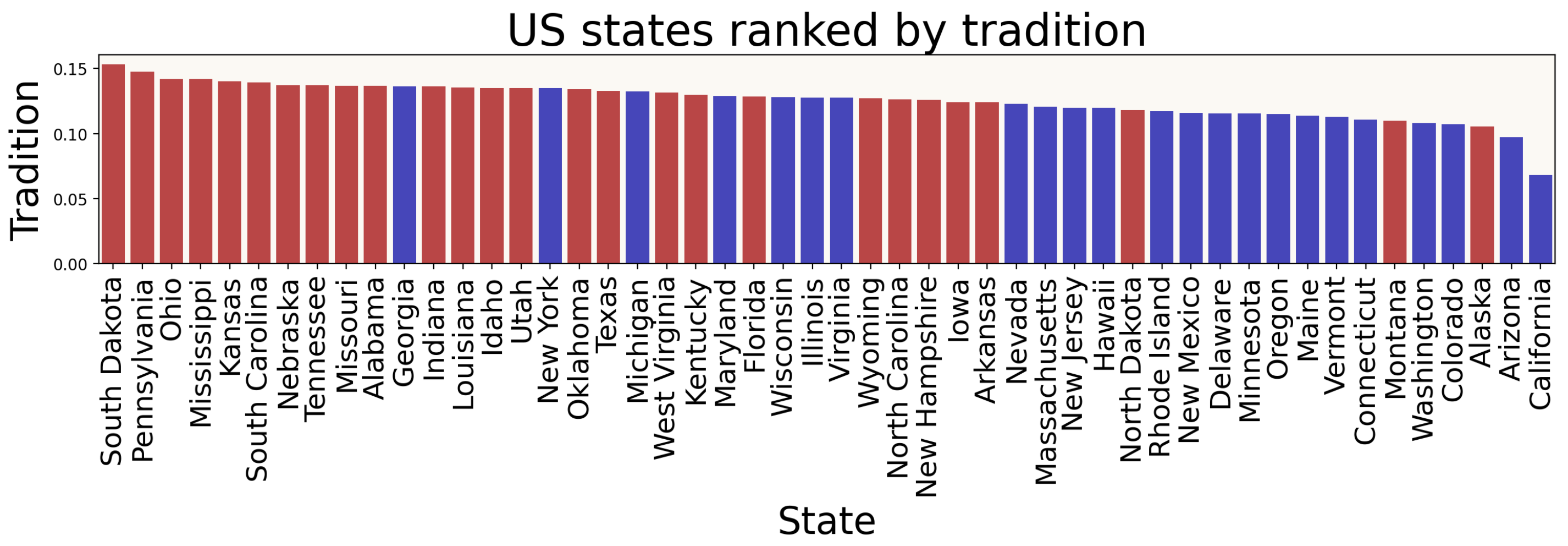
Relevance

$$u^{k}_{\text{stance}}(\mathcal{S}) = \frac{1}{|\mathcal{S}^{k}|}\sum_{i\in\mathcal{S}^{k}} u^{k}_{\text{stance}}(c_i)$$

Stance

Analysis

Alignment with surveys

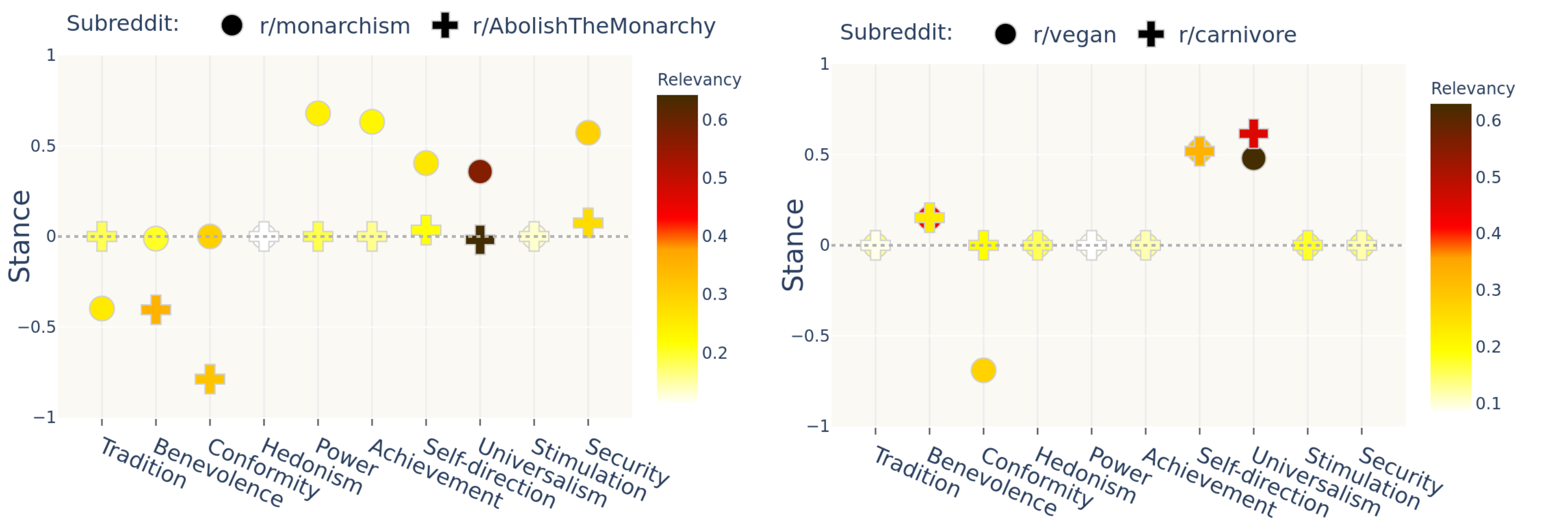


- Good alignment between *tradition* value of state subreddits and surveys.
- No alignment between Schwartz values of country subreddits and value surveys.

Both known and unknown results

Magnitude	Subreddits
Maximal	r/DebateAnarchism, r/Abortiondebate, r/therapyabuse, r/CapitalismVSocialism, r/changemyview, r/AvoidantAttachment, r/LeftWingMaleAdvocates, r/coparenting, r/ADHD_partners, r/DebateAVegan, r/Ask_Politics, r/IsraelPalestine, r/PoliticalDiscussion, r/AskSocialScience, r/NarcAbuseAndDivorce, r/AskDID, r/attachment_theory, r/Adoption, r/kpoprants, r/TrueUnpopularOpinion
Minimal	r/vegan1200isplenty, r/caloriecount, r/Watchexchange, r/Brogress, r/crystalgrowing, r/sneakermarket, r/gundeals, r/buildapcsales, r/whatisit, r/NMSCoordinateExchange, r/BulkOrCut, r/astrophotography, r/legodeal, r/whatisthisthing, r/whatsthisfish, r/filmfashion, r/TipOfMyFork, r/1500isplenty, r/safe_food, r/Repbudgetfashion

Controversial topics



Qualitative analysis

Value	Positive Stance	Negative Stance
Tradition	r/Ankrofficial, r/lds, r/CharliDameLioMommy, r/Christian, r/AskAPriest, r/Bible, r/bahai, r/Quakers, r/PrismaticLightChurch, r/OrthodoxChristianity	r/SuperModelIndia, r/Jewdank, r/EX-HINDU, r/DesiMeta, r/linguisticshumor, r/exmuslim, r/AsABlackMan, r/Satan, r/IndoEuropean, r/AfterTheEndFanFork
Benevolence	r/freebsd, r/RandomKindness, r/Teraform, r/Petloss, r/nextjs, r/Wetshaving, r/AllCryptoBets, r/NixOS, r/vancouverhiking, r/ansible	r/FromDuvalToDade, r/CrimeInTheD, r/NBAYoungboy, r/40kOrkScience, r/LILUZIVERTLEAKS, r/DuvalCounty, r/Phillyscoreboard, r/Chiraqhits, r/SummrsXo, r/CARTILEAKS
Conformity	r/Ankrofficial, r/nanotrade, r/Nervos-	r/Animewallpaper, r/kencarson, r/From-

Conclusions

- We train LMs to predict human values in social media communities.
- Human values are highly subjective, leading to noise in annotations and predictions.
- Overcoming this noise by studying entire populations and not individual posts.
- We uncover both known phenomena and novel insights.

Limitations

- Human values are inherently subjective, leading to unavoidable uncertainty and noise.
- Aggregating subreddit values into a single vector simplifies analysis but overlooks individual post-level dynamics.
- Our approach can identify interesting trends but cannot fully explain them.