Multi-modal Computational Methods in Political Science

Multi-modal Computational Analysis in Political Science

Clara Fochler

Ludwig-Maximilians-Universität München

July 14, 2026

LMU LUDWIG-MAXIMILIANS-UNIVERSITÄT MÜNCHEN

Agenda

Estimating the Ideology of Political YouTube Videos

A multi-modal approach for content detection in Tamil

LMU LUDWIG-MAXIMILIANS-UNIVERSITÄT MÜNCHEN

Lai et al. (2024): Estimating the Ideology of Political YouTube Videos

Research Gap

- YouTube among the most-visited platforms worldwide - yet almost no scalable measure of video-level ideology
- Existing datasets label ideology only at the channel level, based on costly, static human coding
- Channel-level labels hide within-channel variation and can not keep pace with a constantly shifting platform
- **Guiding question:** how can we estimate ideology at the video level?

LMU LUDWIG-MAXIMILIANS-UNIVERSITÄT MÜNCHEN

Preliminary Literature

- Classic ideology estimation focused on traditional political actors: legislators via roll-call votes (Poole & Rosenthal 1985), donors via contributions (Bonica 2014), parties via text scaling (Slapin & Proksch 2008)
- More recent work extends these methods to ordinary social media users (Barberá 2015; Bond & Messing 2015) and to news domains/websites (Robertson et al. 2018; Eady et al. 2020)
- YouTube-specific studies rely on small, manually labeled channel-level data sets (Hosseinmardi et al. 2020; Munger & Phillips 2022; Ribeiro et al. 2020) - static, labor-intensive, and unable to capture video-level or within-channel variation
- Common thread across this literature: the **homophily assumption** - actors and the things they choose or support tend to share similar ideological positions

LMU LUDWIG-MAXIMILIANS-UNIVERSITÄT MÜNCHEN

Data Collection

- Basis: Pushshift Reddit dataset, all posts linking to YouTube videos (Dec. 2011–June 2021), $\sim$31 million posts
- Subreddit identification
  - Seed set of 424 hand-picked political subreddits
  - Expanded through subreddit co-posting network + Leiden community detection
  - Political vs. non-political subreddits separated via correspondence analysis
  - Final set: 819 political subreddits
- Filtering: subreddits with $\geq$ 5 unique videos, videos posted in $\geq$ 3 subreddits $\rightarrow$ 432,115 posts
- Resulting matrix: 74,038 videos $\times$ 685 subreddits, cell entries = log(post score+1)
- Validation data: watch histories of 3,337 YouGov respondents (Aslett et al. 2022); additional human-coded video labels

LMU LUDWIG-MAXIMILIANS-UNIVERSITÄT MÜNCHEN

Method: Two-Step Estimation

Step 1 - Correspondence Analysis

- Standardize the subreddit-video matrix, apply singular value decomposition
- First dimension of the video (row) solution $\rightarrow$ ideology score X_i
- First dimension of the subreddit (column) solution $\rightarrow$ ideology score θ_j, used for validation
- Ideology estimates obtained for 61,883 videos with available text metadata

Step 2 - Text-Based Prediction (BERT)

- Fine-tune BERT on title, channel name, description, and tags to predict the Step-1 score
- Training set 49,970 videos, validation 2,631, test 9,282
- Allows ideology estimation for *any* political video, not just those posted to Reddit

LMU LUDWIG-MAXIMILIANS-UNIVERSITÄT MÜNCHEN

Results: Model Performance & Validation

- Text model tracks the matrix-based scores well: correlation 0.891 ($R^2 \approx 0.793$), MAE ≈ 0.295
- Face validity: subreddit ordering matches domain expectations (e.g. Sanders/Yang communities on the left, r/The_Donald and far-right subreddits on the right)
- Channel-level scores align with prior five-category human-coded labels (Hosseinmardi et al. 2020); IQR narrows toward the ideological extremes
- Human coders: agreement rises above 75% once score distance exceeds 1; correlation of 0.66 with human bin placements

LMU LUDWIG-MAXIMILIANS-UNIVERSITÄT MÜNCHEN

Results: Applications

Media diets

- Median video ideology by self-reported party: Democrats −0.36, Independents 0.26, Republicans 0.53
- Within-respondent spread narrower than the overall distribution – suggests selective, congruent viewing

Ideology and engagement

- Ideologically extreme videos draw disproportionately more likes/dislikes and comments per view
- No clear relationship between ideology and raw view counts
- Points to a possible mechanism by which engagement-based recommendation systems favor extreme content

LMU LUDWIG-MAXIMILIANS-UNIVERSITÄT MÜNCHEN

Mohan et al. (2025): A multi-modal approach for hate and offensive content detection in Tamil

Research Gap

- research on detecting hate speech and offensive language focuses on selective languages - mainly English
- social media more and more multi-modal → analyzing multi-modal is challenging because ...

  ... difficult to get features from different modalities and train models based on it

  ... different length & quality of multi-modal content across (and within) platforms

  ... analyzing different types of hate content difficult due to language nuances & words with multiple meanings (same for symbols)

  ... this is ESPECIALLY the case for languages such as Tamil, where there is no extensive gold standard dataset available

LMU LUDWIG-MAXIMILIANS-UNIVERSITÄT MÜNCHEN

Related Work: multi-modal Sentiment & Emotion Datasets

- Zadeh et al. built **MOSI**, over 2,000 sentiment, intensity, valence and arousal annotated YouTube videos with diverese content (including political speeches); Liang et al. later reached an F1 of 0.70 on it with their Recurrent Multistage Fusion Network model
- Busso et al.'s **Interactive Emotional Dyadic Motion Capture** recorded 151 acted dialogues (302 speakers) labelled with nine emotions; Tripathi et al. combined CNN (visual input) and LSTM (speech & text inout) branches on it and got 82% accuracy
- Poria et al.'s **MELD** dataset contains annotated by emotion "naturalistic conversations" in laboratory settings between 3 people also utilizing*Friends TV series* dialogues; Song et al.'s Supervised Prototypical Contrastive Learning approach (SPCL) used this dataset
- **CMU-MOSEI** BY (Zadeh et al.) with over 23,000 clips, 40,000 sentences, and comes with their own IDFG model for interpretable fusion
- Hasan et al.'s **UR-FUNNY** takes a different angle, using comedic clips to study humour recognition rather than sentiment or emotion directly

LMU LUDWIG-MAXIMILIANS-UNIVERSITÄT MÜNCHEN

Related Work: Dravidian & Tamil Datasets

- Chakravarthi et al. were among the first to release code-mixed Tamil and Malayalam data for offensive language detection, drawing from social media comments
- Roy et al. approached the same problem with an ensemble of CNNs, LSTMs and transformer models (BERT, MuRIL), comparing this against classical TF-IDF baselines
- Subramanian et al. tested both traditional classifiers and transformer models on Tamil offensive language, and found that adapter-based fine-tuning of mBERT, MuRIL and XLM-R gave the best results
- the one exception that actually goes multi-modal is Chakravarthi et al.'s sentiment dataset built from 134 Tamil and Malayalam YouTube review videos
- what's still missing, and the paper wants to add: is a multi-modal dataset for *hate speech* specifically in Tamil

LMU LUDWIG-MAXIMILIANS-UNIVERSITÄT MÜNCHEN

The MATH Dataset: Data Collection

- Tamil is the oldest Dravidian language, spoken in some parts of India, Singapore, Malaysia etc.
- **52 YouTube videos** collected and annotated, each between **1 and 3 minutes** long
- Three modalities extracted per video: **text** (transcribed speech), **audio**, and **video**
- Videos labelled across four categories: **offensive**, **racist**, **sexist**, and **casteist** content
- Class distribution is imbalanced: offensive (17) and racist (16) dominate, sexist is rare (4)

LMU LUDWIG-MAXIMILIANS-UNIVERSITÄT MÜNCHEN

Challenges in Building a Tamil Hate Speech Dataset

- **Code-mixing**: Tamil frequently mixed with English or other regional languages, complicating detection
- **Audio quality**: background noise and low-quality recordings hindered transcription
- **Limited captions**: Tamil subtitles are rare, and automated speech-to-text tools performed poorly
- **Dialect variation**: multiple regional dialects made consistent annotation difficult
- As a result, transcripts had to be created **manually** rather than relying on IBM/Google speech-to-text

LMU LUDWIG-MAXIMILIANS-UNIVERSITÄT MÜNCHEN

Annotation Process

- Annotation followed procedures adapted from Zadeh et al. and Mohammad
- **Three independent annotators**, all Tamil-fluent postgraduates, labelled each clip separately to reduce bias
- Labels assigned using all three modalities together: **facial expressions**, **vocal tone**, and **transcript content**
- Particularly useful for detecting **sarcasm**, where tone and expression diverge from the literal text

LMU LUDWIG-MAXIMILIANS-UNIVERSITÄT MÜNCHEN

Ablation Study: Unimodal Performance

- **Text**: four transformer models tested (Tamil-BERT, MuRIL-Large, LaBSE, Hate-MuRIL); **Tamil-BERT + Logistic Regression** performed best at **81.82% accuracy**
- **Audio**: features extracted with **Wav2Vec2**; removing background noise (spectral subtraction) improved results to **72.72% accuracy**
- **Video**: **TimeSformer** used to capture spatiotemporal cues, outperforming an Inception Net baseline, though accuracy stayed comparatively low ($\sim$45%)
- Inter-annotator agreement (Cohen's Kappa) reached **0.72**, indicating substantial agreement among the three annotators

LMU LUDWIG-MAXIMILIANS-UNIVERSITÄT MÜNCHEN

Multi-modal Fusion & Conclusion

- Best fusion combined **Hate-MuRIL** (text) → **Noise-removed Wav2Vec2** (audio) → **TimeSformer** (video)
- With Logistic Regression as classifier, this combination reached **81.82% accuracy**, matching the best unimodal (text-only) result – what is the issue with taking accuracy as performance measure here?
- Text and audio modalities contributed most to classification; video remained the weakest signal
- **MATH** is the first multi-modal hate speech dataset for Tamil, filling a clear gap for low-resource language research
- Future work: stronger video modality techniques and more advanced multi-modal fusion strategies

LMU LUDWIG-MAXIMILIANS-UNIVERSITÄT MÜNCHEN

REFERENCES

- Lai, A., Brown, M. A., Bisbee, J., Tucker, J. A., Nagler, J., Bonneau, R. (2024). Estimating the ideology of political YouTube videos. Political Analysis, 32(3), 345–360. https://doi.org/10.1017/pan.2023.42
- Mohan, J., Mekapati, S. R., Premjith, B., Lal G, J., Chakravarthi, B. R. (2025). A multi-modal approach for hate and offensive content detection in Tamil: From corpus creation to model development. ACM Transactions on Asian and Low-Resource Language Information Processing, 24(3), Article 22. https://doi.org/10.1145/3712260

LMU LUDWIG-MAXIMILIANS-UNIVERSITÄT MÜNCHEN

Thanks for your attention!