Gauging the Behavioral Patterns of Voters Using a Unified Theory of Social Media Modeling

Tushaar Gangavarapu (BTech)
Prof. Dr. Ram Mohana Reddy Guddeti (PhD)

Introduction

- ❖ Elections provide important opportunity to advance democratization!
- ❖ Now-a-days social media stands as an effective platform of self-expression, communication, and participation
    - ❖ Pitfalls of social media data*
    - ❖ Bots and sockpuppets?
    - ❖ Escaping filter bubbles?
    - ❖ Linguistic and temporal structure

* Kosinski, M., et al. "Facebook as a Research Tool for the Social Sciences." American Psychologist (2015): 543-556.

Introduction

- Elections provide important opportunity to advance democratization!
- Now-a-days social media stands as an effective platform of self-expression, communication, and participation
- India: 2013-14 was the first exposure of social media for electoral campaigning
- Time and cost effectiveness of using social media to gauge public opinion

Introduction

- Elections provide important opportunity to advance democratization!
- Now-a-days social media stands as an effective platform of self-expression, communication, and participation
- India: 2013-14 was the first exposure of social media for electoral campaigning
- Time and cost effectiveness of using social media to gauge public opinion

- Can this carefully judged insight ever replace the traditional polling?

Social Media: Open Source to Privacy?

- Finding patterns and trends with time
  - Sources
  - Sentiments
  - Hashtags
  - Alliances

Social Media: Open Source to Privacy?

- Finding patterns and trends with time
  - Sources
  - Sentiments
  - Hashtags
  - Alliances
- Sustainability and popularity on the social network
  - Temporal nature
  - User mentions

Social Media: Open Source to Privacy?

State	Political Party	Results
Telangana	BJP	1/119
	Congress+	21/119
	TRS	88/119
	Others	9/119
Chhattisgarh	BJP	22/90
	Congress+	68/90
Rajasthan	BJP	79/199
	Congress+	100/199
	Others	20/199
Madhya Pradesh	BJP	111/230
	Congress+	114/230
	Others	5/230
Mizoram	MLF	26/40
	Others	14/40

Opinion Polling: India 2018-19

- General elections for 5 states – Telangana, Chhattisgarh, Rajasthan, Madhya Pradesh, and Mizoram

Opinion Polling: India 2018-19

- General elections for 5 states – Telangana, Chhattisgarh, Rajasthan, Madhya Pradesh, and Mizoram
- BJP vs. Congress at the “national level” with various alliances at the “state level”

Opinion Polling: India 2018-19

- General elections for 5 states – Telangana, Chhattisgarh, Rajasthan, Madhya Pradesh, and Mizoram
- BJP vs. Congress at the "national level" with various alliances at the "state level"
- Legacy data on previously formed alliances can help simulate post-election possible alliances (monte-carlo simulations)

Opinion Polling: India 2018-19

- General elections for 5 states – Telangana, Chhattisgarh, Rajasthan, Madhya Pradesh, and Mizoram
- BJP vs. Congress at the "national level" with various alliances at the "state level"
- Legacy data on previously formed alliances can help simulate post-election possible alliances (monte-carlo simulations)
- Indian elections in comparison with U.S. elections have many unmodelable "influential parameters" (e.g., money, liquor, gifts, schemes, etc.)

Opinion Polling: India 2018-19

- General elections for 5 states – Telangana, Chhattisgarh, Rajasthan, Madhya Pradesh, and Mizoram
- BJP vs. Congress at the "national level" with various alliances at the "state level"
- Legacy data on previously formed alliances can help simulate post-election possible alliances (monte-carlo simulations)
- Indian elections in comparison with U.S. elections have many unmodelable "influential parameters" (e.g., money, liquor, gifts, schemes, etc.)
- Indian election prediction involves mining the opinions of people as well as collection of data from the official handles

Opinion Polling: India 2018-19

- General elections for 5 states – Telangana, Chhattisgarh, Rajasthan, Madhya Pradesh, and Mizoram
- BJP vs. Congress at the "national level" with various alliances at the "state level"
- Legacy data on previously formed alliances can help simulate post-election possible alliances (monte-carlo simulations)
- Indian elections in comparison with U.S. elections have many unmodelable "influential parameters" (e.g., money, liquor, gifts, schemes, etc.)
- Indian election prediction involves mining the opinions of people as well as collection of data from the official handles
- Various polling strategies (e.g., open debates, allegations) must also be analyzed

Outcome of Literature Survey

- Sampling bias problem – not all people registered to vote express self-opinions on social media; geo-tagging can mitigate, not eradicate!

Outcome of Literature Survey

- Sampling bias problem – not all people registered to vote express self-opinions on social media; geo-tagging can mitigate, not eradicate!
- Diversity among modeling – which of the existing state-of-the-art can best model social media data?

Outcome of Literature Survey

- Sampling bias problem – not all people registered to vote express self-opinions on social media; geo-tagging can mitigate, not eradicate!
- Diversity among modeling – which of the existing state-of-the-art can best model social media data?
- Temporal nature of the social media – sustaining preference for a political party over a long period of time vs. the rising popularity of an event?

Outcome of Literature Survey

- Sampling bias problem – not all people registered to vote express self-opinions on social media; geo-tagging can mitigate, not eradicate!
- Diversity among modeling – which of the existing state-of-the-art can best model social media data?
- Temporal nature of the social media – sustaining preference for a political party over a long period of time vs. the rising popularity of an event?
- Affective outcomes of party choices – influence of the events caused by an electoral candidate or a political party

Outcome of Literature Survey

- Sampling bias problem – not all people registered to vote express self-opinions on social media; geo-tagging can mitigate, not eradicate!
- Diversity among modeling – which of the existing state-of-the-art can best model social media data?
- Temporal nature of the social media – sustaining preference for a political party over a long period of time vs. the rising popularity of an event?
- Affective outcomes of party choices – influence of the events caused by an electoral candidate or a political party
- Gender-based analytics – if social media mimics the real-world scenario, then should gender-based analytics be incorporated into the prediction model?

Problem Statement and Research Objectives

Mining and modeling users' digital footprints on social media (Twitter) to accurately predict the outcomes of the general elections in a nation (India)

Research Objectives:

- Design of an effective modeling framework that unifies various theories of social media modeling including volumetric, sentiment, network, and gender analyses to gauge at voters' behavioral patterns

Problem Statement and Research Objectives

Mining and modeling users' digital footprints on social media (Twitter) to accurately predict the outcomes of the general elections in a nation (India)

Research Objectives:

- Design of an effective modeling framework that unifies various theories of social media modeling including volumetric, sentiment, network, and gender analyses to gauge at voters' behavioral patterns
- Incorporating the temporal nature of social media footprints to assess the sustainability and popularity of a political party concerning day-to-day political changes over a long period

Problem Statement and Research Objectives

Mining and modeling users' digital footprints on social media (Twitter) to accurately predict the outcomes of the general elections in a nation (India)

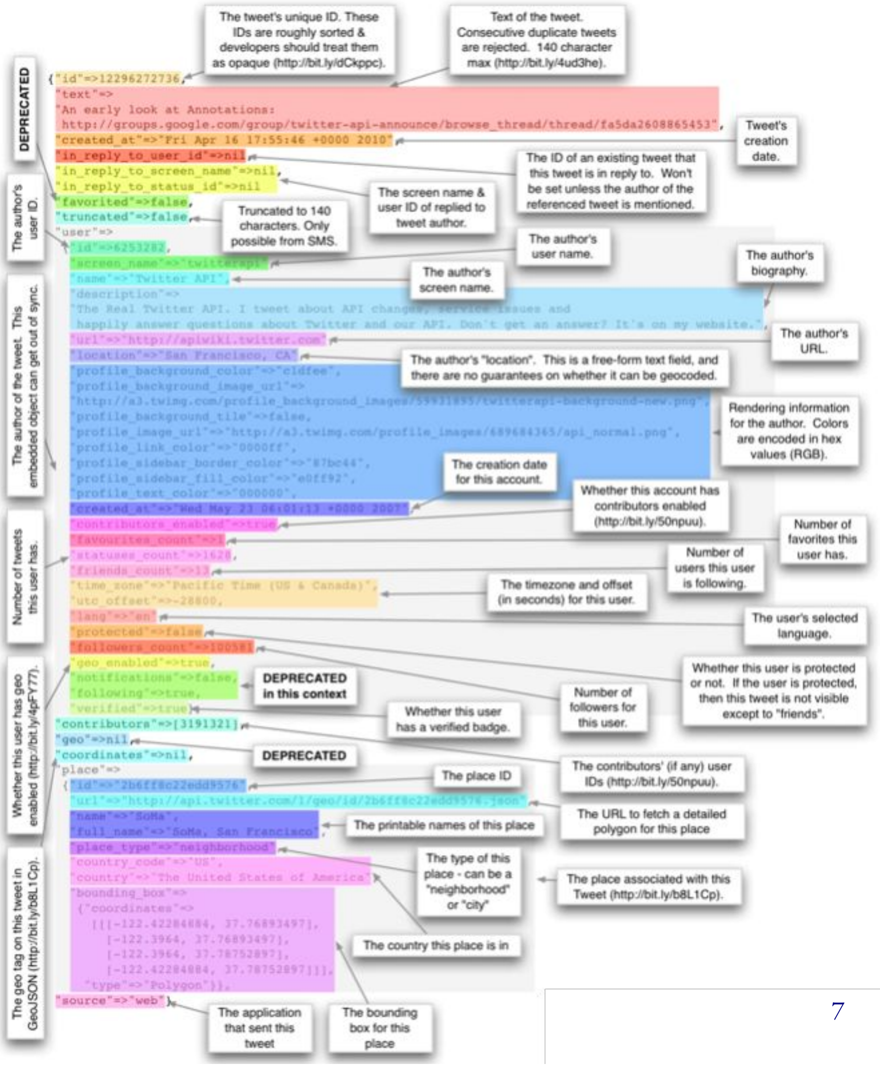
Research Objectives:

- Design of an effective modeling framework that unifies various theories of social media modeling including volumetric, sentiment, network, and gender analyses to gauge at voters' behavioral patterns
- Incorporating the temporal nature of social media footprints to assess the sustainability and popularity of a political party concerning day-to-day political changes over a long period
- Mining the affective outcomes of the voters as the digital footprints of the events caused by an electoral candidate or political party

Twitter Mining: Parameters

- Followers count
- Friends count
- Listed count
- Retweet, Quote count
- Reply, Favorite count

reachability

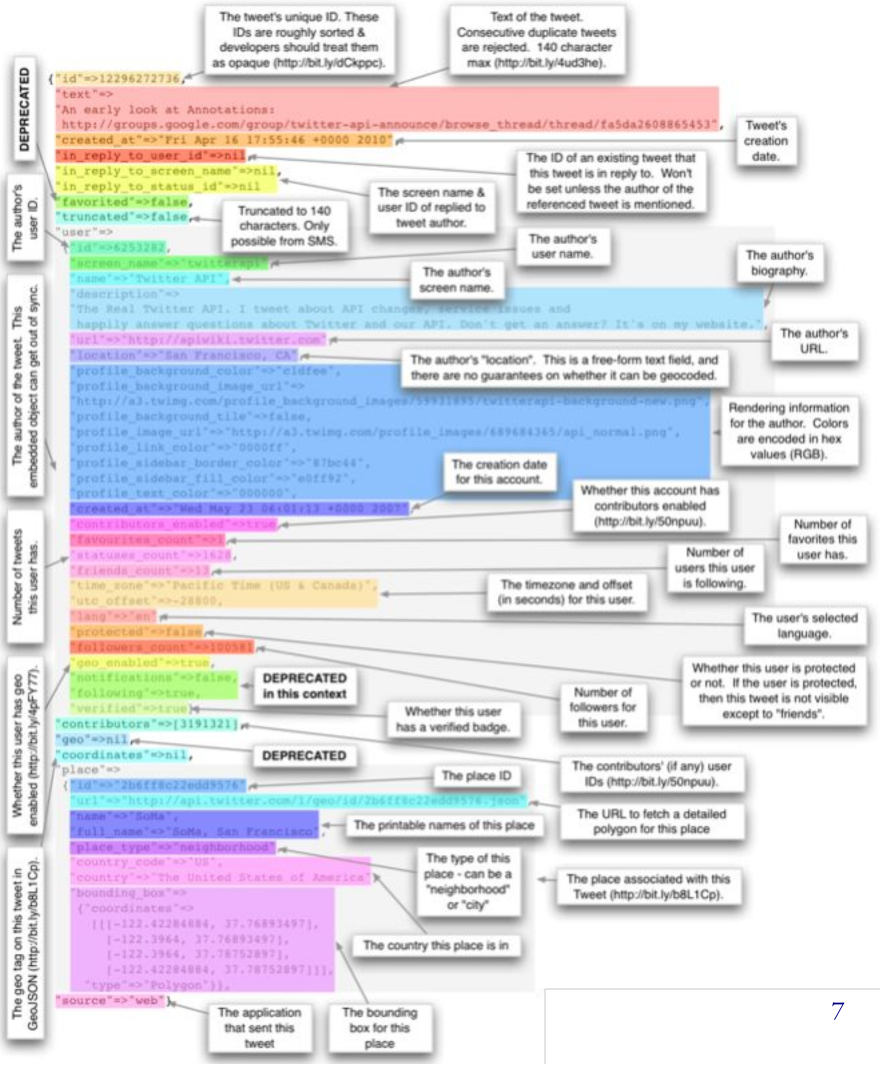


Twitter Mining: Parameters

- Followers count
- Friends count
- Listed count
- Retweet, Quote count
- Reply, Favorite count

} reachability

- Location – geo-tagging
- Time of creation – temporal ranking
- Hashtags, User mentions– network
- Language, Tweet text– sentiment

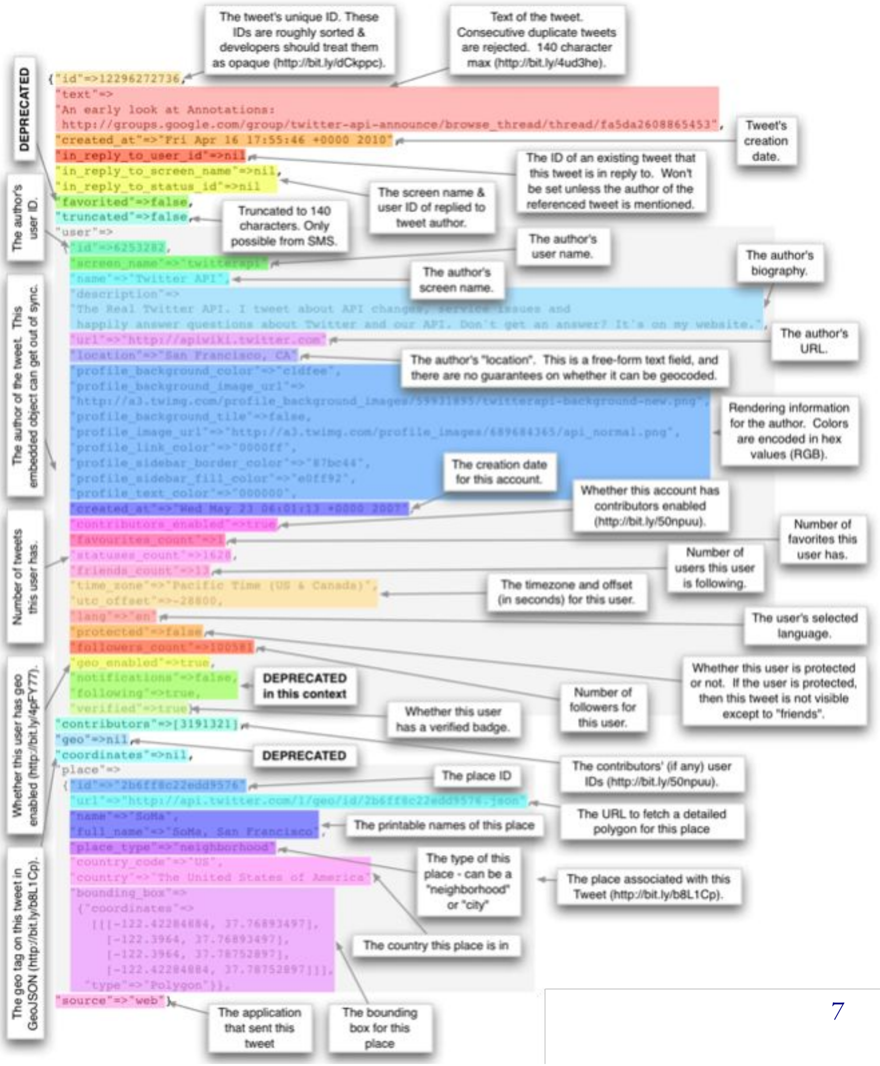


Twitter Mining: Parameters

- Followers count
- Friends count
- Listed count
- Retweet, Quote count
- Reply, Favorite count

} reachability

- Location – geo-tagging
- Time of creation – temporal ranking
- Hashtags, User mentions– network
- Language, Tweet text– sentiment
- Gender* – analysis



* Gender is not a Twitter parameter, it has to estimated separately

Challenges in Twitter Modeling

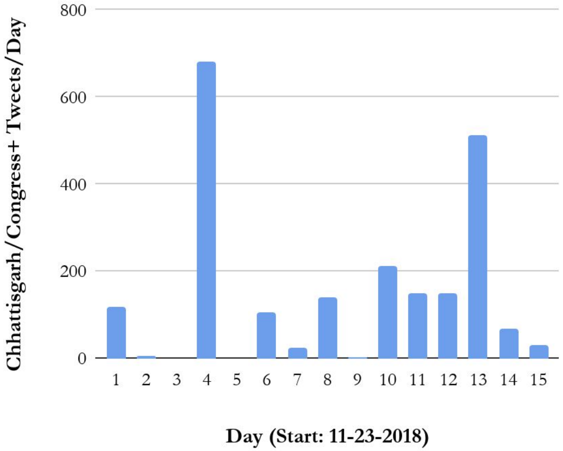
- Location – number of people posting their opinions vs. number of those people voting!

user_screen_name	user_gender	user_followers_count	user_friends_count	user_listed_count	user_location	tweet_created_at
urskumar9	female	1012	233	3	None	Fri Nov 09 04:26:11 +0000 2018
saiverameshwar		4	26	0	None	Fri Nov 09 04:26:30 +0000 2018
tiny_jerk		3	155	0	Milky_way	Fri Nov 09 04:26:48 +0000 2018

≅ 65.07% tweets – unrelated geo-tags

Challenges in Twitter Modeling

- Location – number of people posting their opinions vs. number of those people voting!



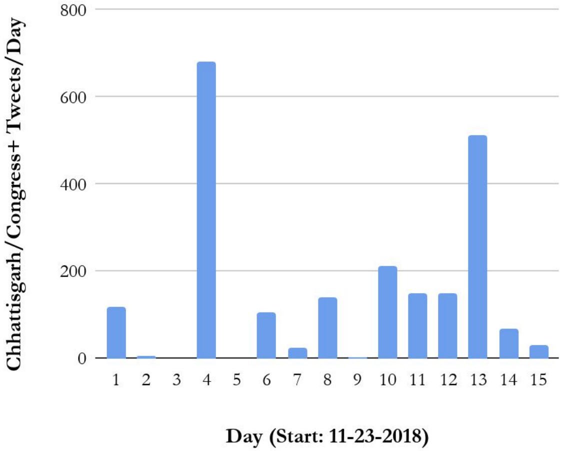
user_screen_name	user_gender	user_followers_count	user_friends_count	user_listed_count	user_location	tweet_created_at
urskumar9	female	1012	233	3	None	Fri Nov 09 04:26:11 +0000 2018
saiverameshwar		4	26	0	None	Fri Nov 09 04:26:30 +0000 2018
tiny_jerk		3	155	0	Milky_way	Fri Nov 09 04:26:48 +0000 2018

≅ 65.07% tweets – unrelated geo-tags

- Volume – not many handles and the number of posts per handle is as low as 3 tweets/day!

Challenges in Twitter Modeling

- Location – number of people posting their opinions vs. number of those people voting!



user_screen_name	user_gender	user_followers_count	user_friends_count	user_listed_count	user_location	tweet_created_at
urskumar9	female	1012	233	3	None	Fri Nov 09 04:26:11 +0000 2018
saiverameshwar		4	26	0	None	Fri Nov 09 04:26:30 +0000 2018
tiny_jerk		3	155	0	Milky_way	Fri Nov 09 04:26:48 +0000 2018

≅ 65.07% tweets – unrelated geo-tags

- Volume – not many handles and the number of posts per handle is as low as 3 tweets/day!
- Language – discarding non-english tweets removes "48.2%" of the tweets (Hindi or Telugu)!

Theories of Social Media Modeling: Volume

- Measures the volume of attention or support, and is computed as the frequency of mentions online (e.g., retweets, supporters, likes, etc.)

$$Vol_{p,t} = \frac{\{\text{relevant social media mentions}\}_{p,t}}{\sum_{i} \{\text{relevant social media mentions}\}_{i,t}} \ (\%)$$

Theories of Social Media Modeling: Volume

- Measures the volume of attention or support, and is computed as the frequency of mentions online (e.g., retweets, supporters, likes, etc.)

$$\mathrm{Vol}_{p,t} = \frac{\{\text{relevant social media mentions}\}_{p,t}}{\sum_i \{\text{relevant social media mentions}\}_{i,t}} \ (\%)$$

- Digital footprints are preprocessed to ensure that the retained footprints specifically mention a single party or its most prominent candidate

Theories of Social Media Modeling: Volume

- Measures the volume of attention or support, and is computed as the frequency of mentions online (e.g., retweets, supporters, likes, etc.)

$$\mathrm{Vol}_{p,t} = \frac{\{\text{relevant social media mentions}\}_{p,t}}{\sum_{i} \{\text{relevant social media mentions}\}_{i,t}} \ (\%)$$

- Digital footprints are preprocessed to ensure that the retained footprints specifically mention a single party or its most prominent candidate
- Volumetric score is temporally weighted to assess the sustainability and popularity with time

$$\mathrm{Tem}_{t,t_e} = \frac{k}{t - t_e}$$

$$\boxed{\mathrm{Vol}^{(t)}_{p,T,t_e} = \frac{\sum_{t} \{\text{relevant social media mentions}\}_{p,t} \cdot \mathrm{Tem}_{t,t_e}}{\sum_{t} \sum_{i} \{\text{relevant social media mentions}\}_{i,t}} \ (\%)}$$

Theories of Social Media Modeling: Sentiment

- Measures the positive, negative, and net sentiment impressions of each party on social media, based on simple counts of the number of tweets with the positive and negative sentiment

Theories of Social Media Modeling: Sentiment

- Measures the positive, negative, and net sentiment impressions of each party on social media, based on simple counts of the number of tweets with the positive and negative sentiment
- Sentiment score ($Sen_{p,T,t_e}^{(t,\ rch)}$) is weighted temporally and by the reachability of the footprint to assess the sustainability and popularity with time
  - User reachability index = *#followers* + *#friends* + *#listed*
  - Tweet reachability index = *#retweets* + *#favorites* + *#replies* + *#quotes*

Theories of Social Media Modeling: Sentiment

- Measures the positive, negative, and net sentiment impressions of each party on social media, based on simple counts of the number of tweets with the positive and negative sentiment
- Sentiment score ($Sen^{(t,\, rch)}_{p,T,t_e}$) is weighted temporally and by the reachability of the footprint to assess the sustainability and popularity with time
  - User reachability index = *#followers* + *#friends* + *#listed*
  - Tweet reachability index = *#retweets* + *#favorites* + *#replies* + *#quotes*
- We employ a hybrid of supervised and unsupervised approaches to effectively estimate the sentiment of the collected footprints
  - English tweets: SentiStrength
  - Non-English tweets: Deep Conv-LSTM architecture

Modeling the Sentiment: Challenges

- Complex linguistic structure
  - English + Native Language: దేశంపట్ల or ప్రజలపట్ల responsibility is not visible
  - Only Native Language: దేశంపట్ల కానీ ప్రజలపట్ల కానీ బాధ్యత కనిపించదు
  - Language written in English: desham patla kani prajala patla kani badhyata kanipinchadu

Modeling the Sentiment: Challenges

- Complex linguistic structure
  - English + Native Language: దేశంపట్ల or ప్రజలపట్ల responsibility is not visible
  - Only Native Language: దేశంపట్ల కానీ ప్రజలపట్ల కానీ బాధ్యత కనిపించదు
  - Language written in English: desham patla kani prajala patla kani badhyata kanipinchadu
- Corpus for sentiment classification
  - Telugu (5,410)[1], Hindi (5,228)[2], and code-mixed (Te-En*: 5,410 and Hi-En[3]: 3,879)

* Śata-Anuvādak partial translation of the obtained Telugu corpus was performed to achieve the Te-En code-mix corpus

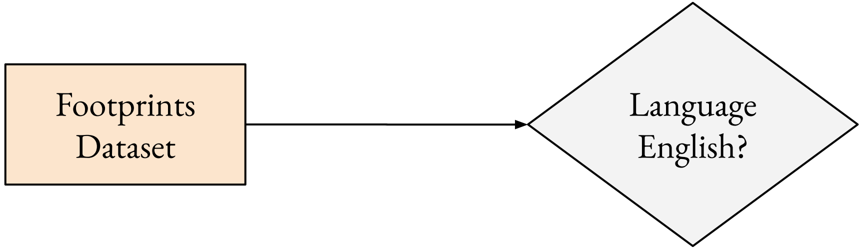
Modeling the Sentiment: Challenges

- Complex linguistic structure
  - English + Native Language: దేశంపట్ల or ప్రజలపట్ల responsibility is not visible
  - Only Native Language: దేశంపట్ల కానీ ప్రజలపట్ల కానీ బాధ్యత కనిపించదు
  - Language written in English: desham patla kani prajala patla kani badhyata kanipinchadu
- Corpus for sentiment classification
  - Telugu (5,410)[1], Hindi (5,228)[2], and code-mixed (Te-En*: 5,410 and Hi-En[3]: 3,879)

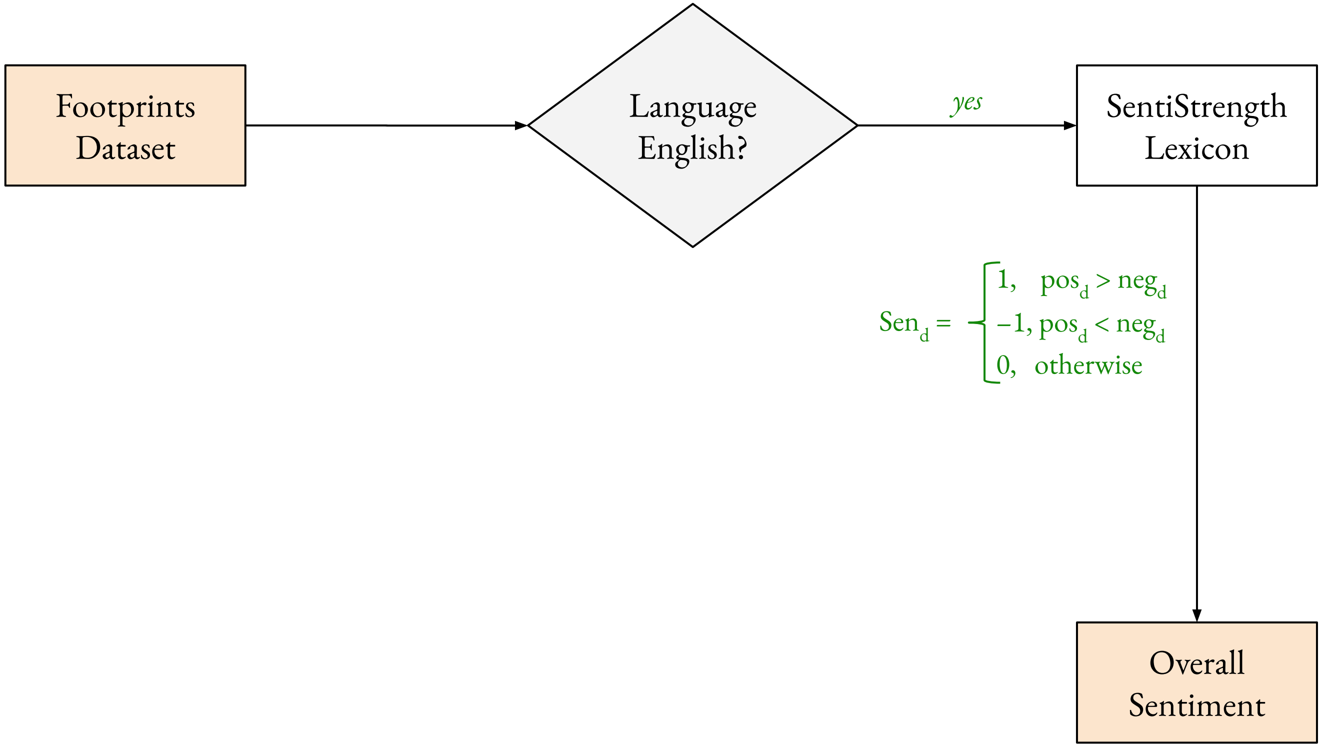
	Telugu			Hindi		
	Positive	Negative	Neutral	Positive	Negative	Neutral
Hand-annotated	250	250	250	250	250	250
Pure language	1,491	1,441	2,478	2,290	712	2,226
English code-mixed	1,491	1,441	2,478	1,352	570	1,957
Total	**3,232**	**3,132**	**5,206**	**3,892**	**1,532**	**4,433**

* Śata-Anuvādak partial translation of the obtained Telugu corpus was performed to achieve the Te-En code-mix corpus

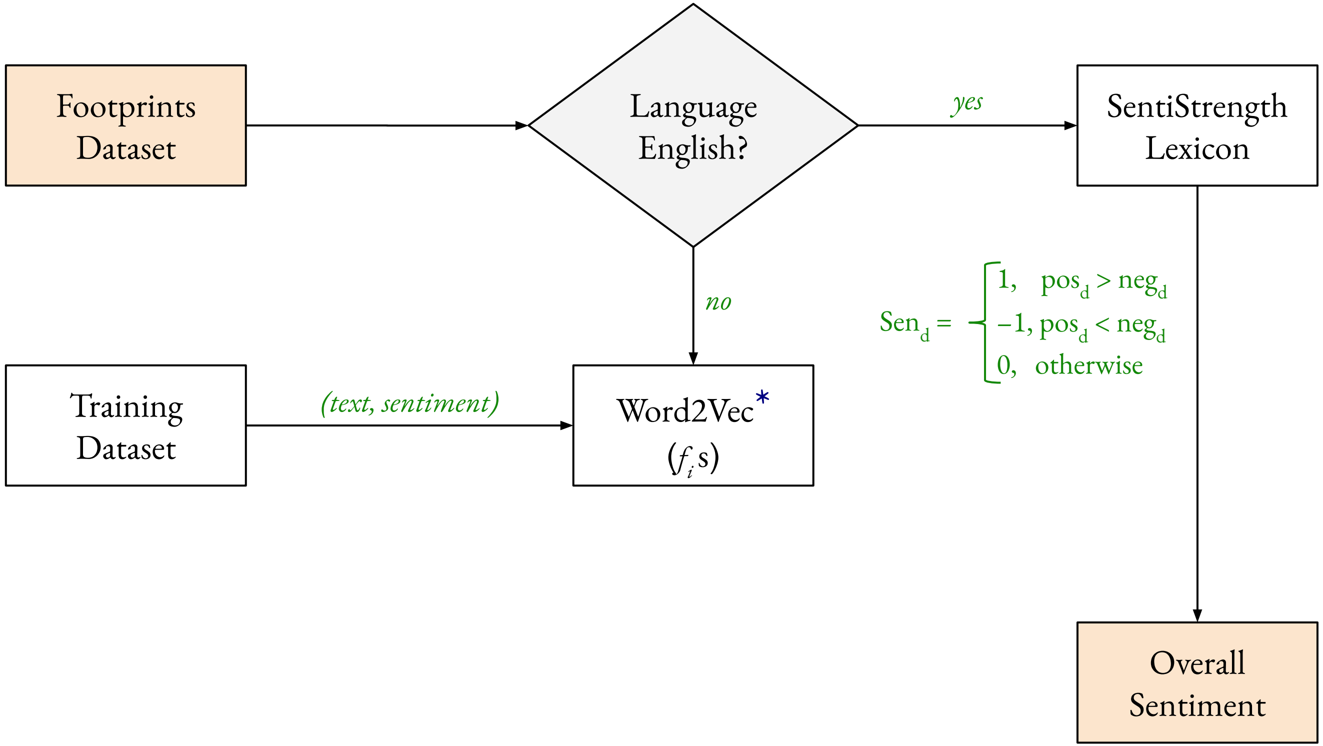
Modeling the Sentiment: Supervised-Unsupervised



Modeling the Sentiment: Supervised-Unsupervised

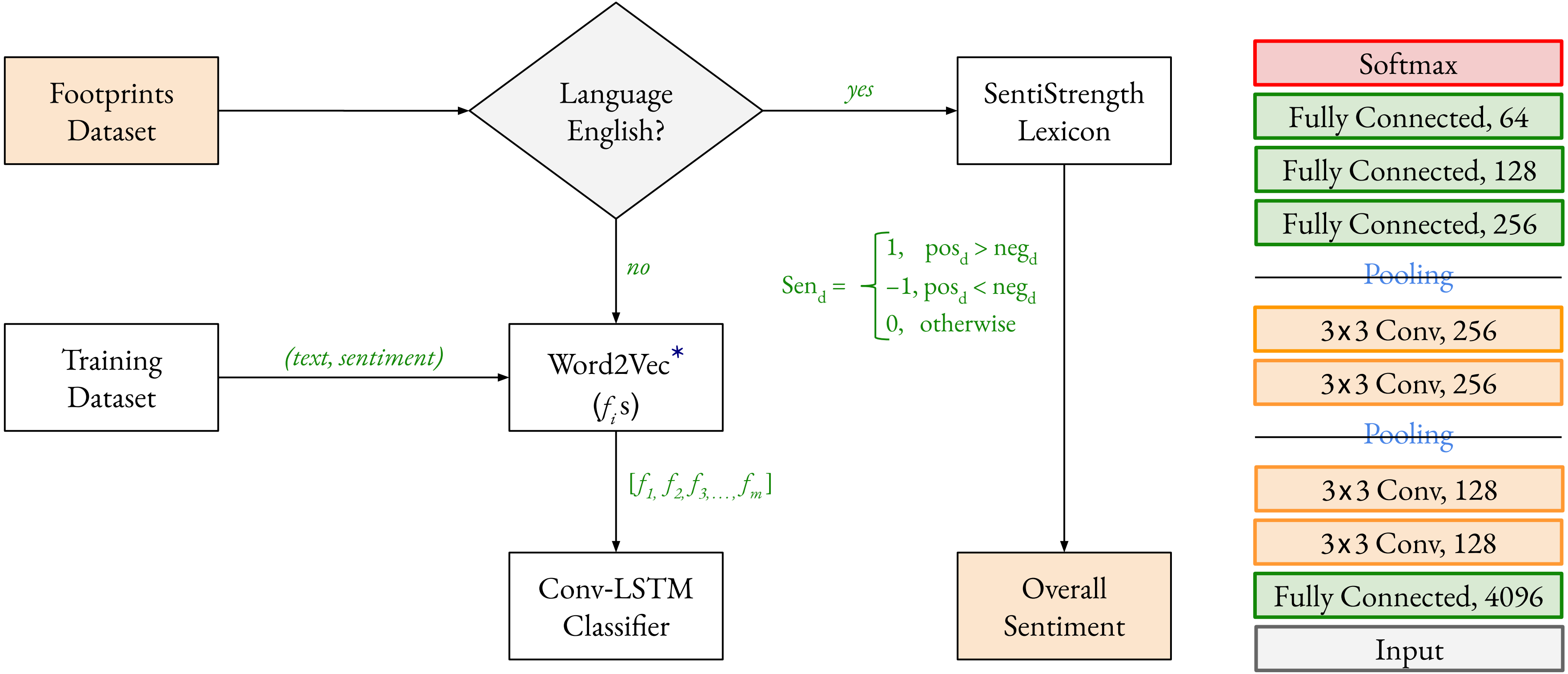


Modeling the Sentiment: Supervised-Unsupervised



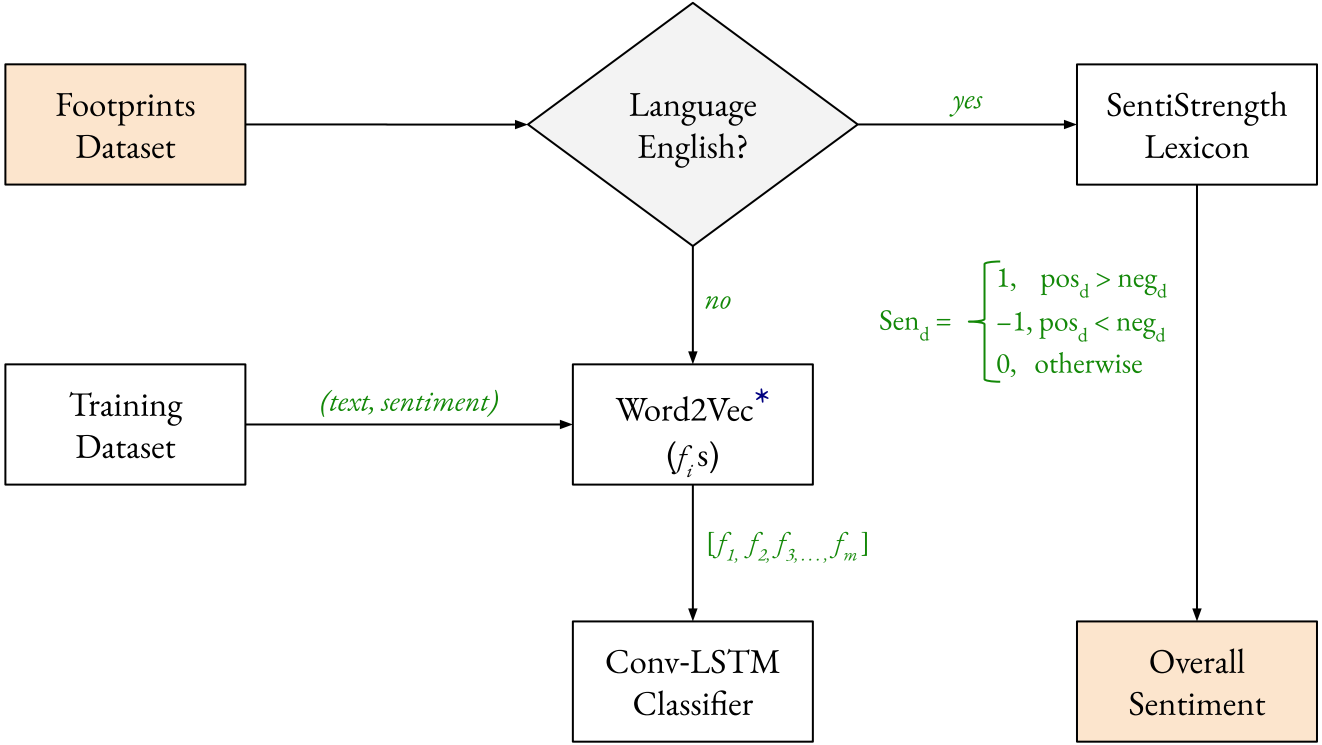
* Pre-trained on 360,000 Hindi and 220,000 Telugu sentences from Wikipedia and Twitter

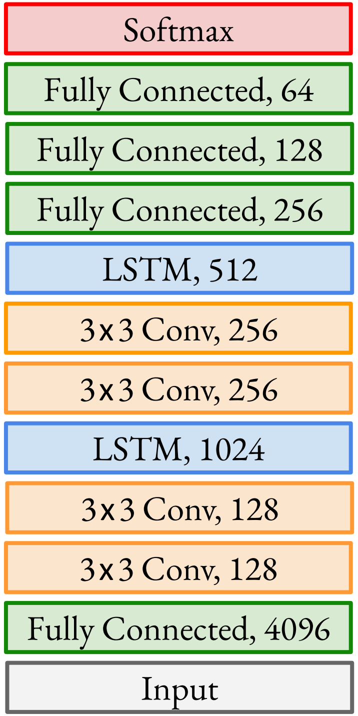
Modeling the Sentiment: Supervised-Unsupervised



* Pre-trained on 360,000 Hindi and 220,000 Telugu sentences from Wikipedia and Twitter

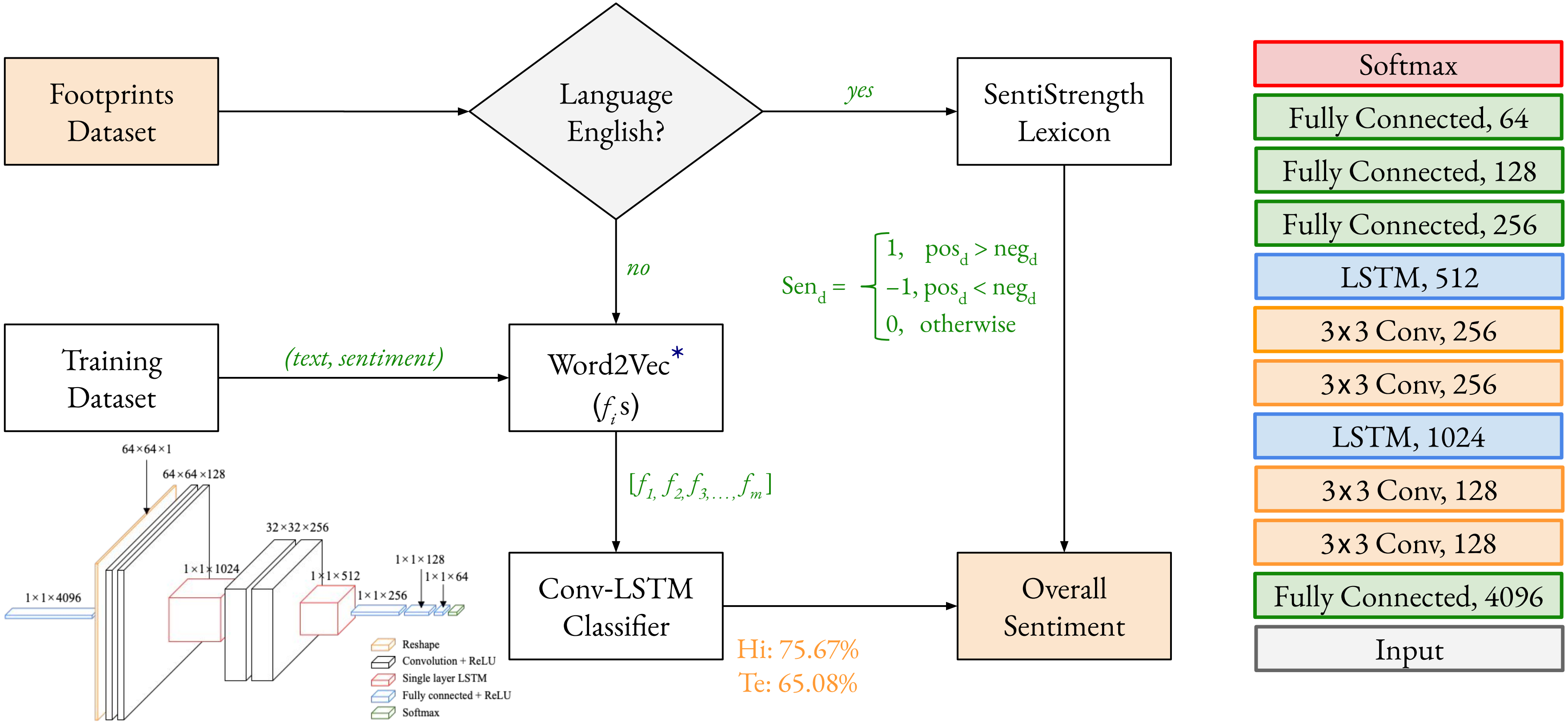
Modeling the Sentiment: Supervised-Unsupervised





* Pre-trained on 360,000 Hindi and 220,000 Telugu sentences from Wikipedia and Twitter

Modeling the Sentiment: Supervised-Unsupervised



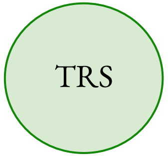
* Pre-trained on 360,000 Hindi and 220,000 Telugu sentences from Wikipedia and Twitter

Theories of Social Media Modeling: Social Network

- Measures the strength of the online community supporting each political party to assess the central position played by the party in the community

Theories of Social Media Modeling: Social Network

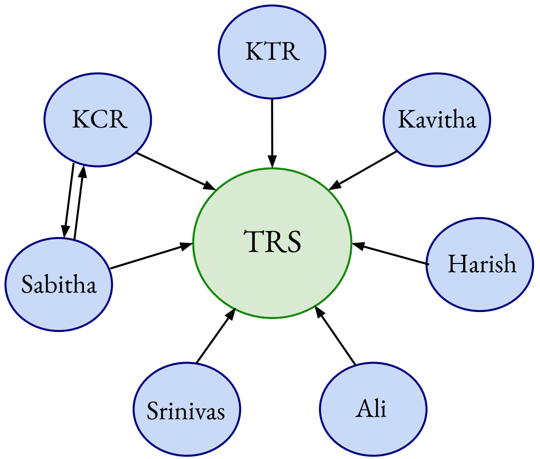
- Measures the strength of the online community supporting each political party to assess the central position played by the party in the community
- Model the social network graph for each political party
    - Central hub: Political party



* Only positively labeled footprints are used

Theories of Social Media Modeling: Social Network

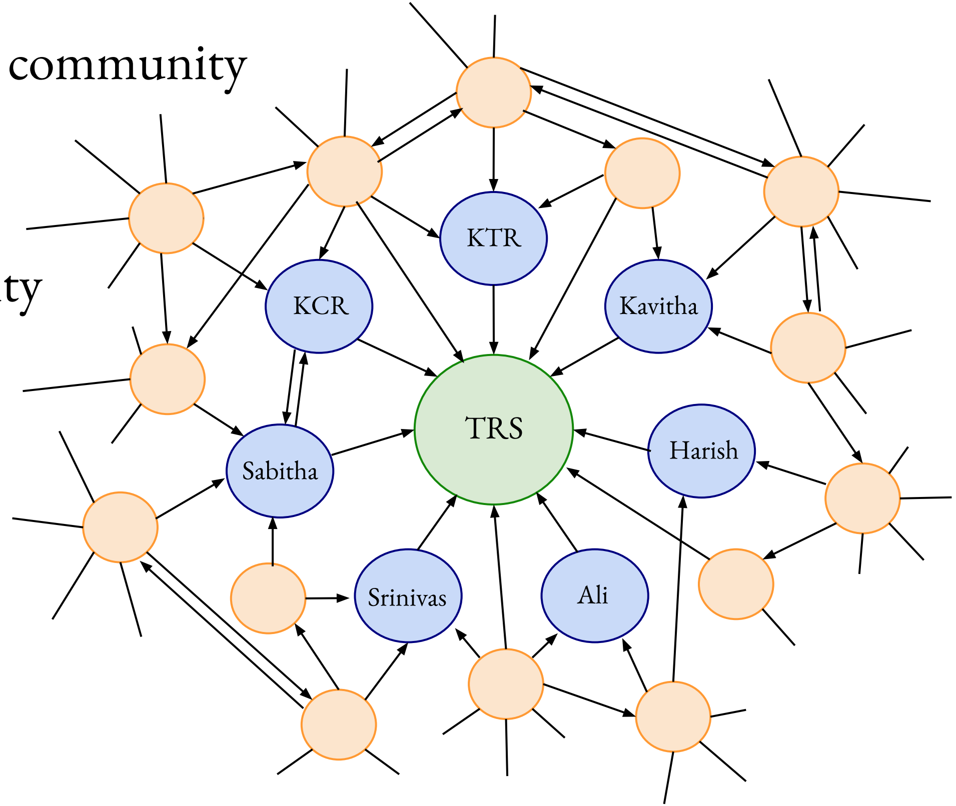
- Measures the strength of the online community supporting each political party to assess the central position played by the party in the community
- Model the social network graph for each political party
  - Central hub: Political party
  - Member nodes: Party candidates



* Only positively labeled footprints are used

Theories of Social Media Modeling: Social Network

- Measures the strength of the online community supporting each political party to assess the central position played by the party in the community
- Model the social network graph for each political party
  - Central hub: Political party
  - Member nodes: Party candidates
  - Other nodes: @/#-mentions



* Only positively labeled footprints are used

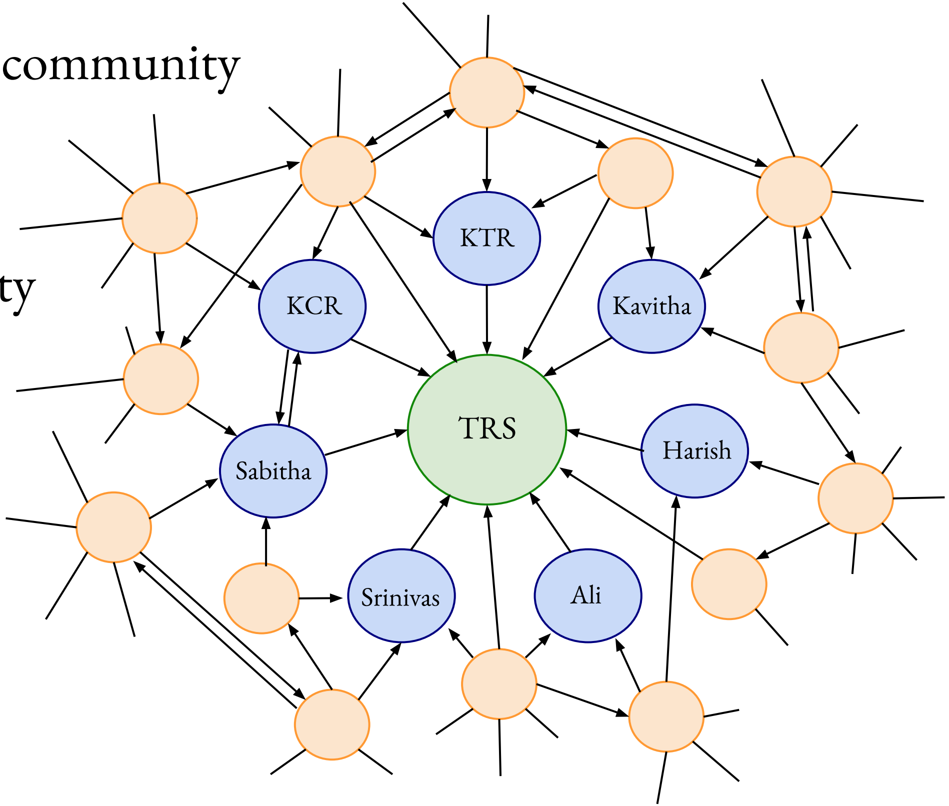
Theories of Social Media Modeling: Social Network

- Measures the strength of the online community supporting each political party to assess the central position played by the party in the community
- Model the social network graph for each political party
- Overall network score:

$$\frac{\#\text{edges}}{\text{possible }\#\text{edges}}$$

$$\boxed{Net^{(dens,\, betw)}_{p,T} = \frac{C_B(p) + (1/D(p))}{\sum_i C_B(i) + (1/D(p))}\ (\%)}$$

$$C_B(p) = \sum_{u \neq v \neq p \in V} \frac{\sigma(u, v \mid p)}{\sigma(u, v)}$$



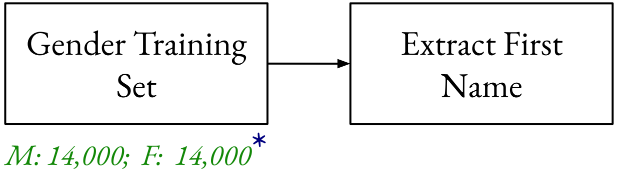
* Only positively labeled footprints are used

Theories of Social Media Modeling: Gender

- Estimates the social media users' gender to closely mimic the real-world scenario
- Estimate the gender gaps and gender stereotyping

Theories of Social Media Modeling: Gender

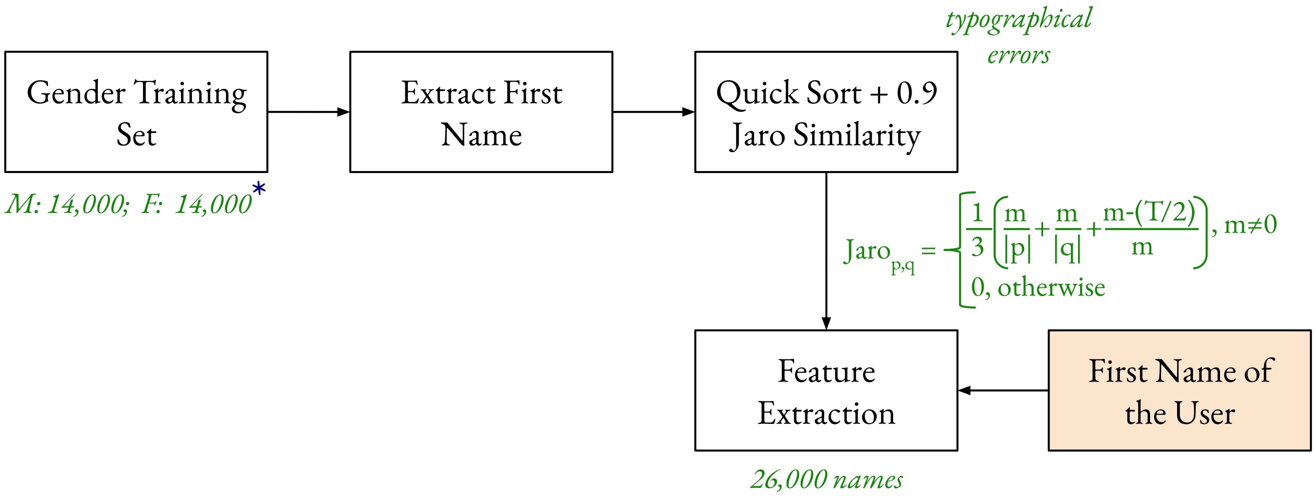
- Estimates the social media users' gender to closely mimic the real-world scenario
- Estimate the gender gaps and gender stereotyping



* M: https://gist.github.com/mbejda/7f86ca901fe41bc14a63; F: https://gist.github.com/mbejda/9b93c7545c9dd93060bd

Theories of Social Media Modeling: Gender

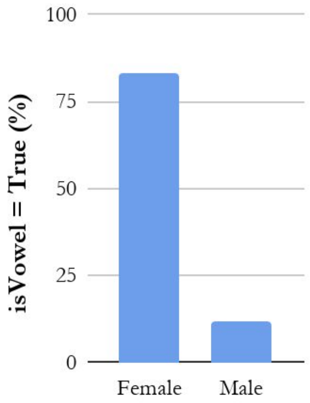
- Estimates the social media users' gender to closely mimic the real-world scenario
- Estimate the gender gaps and gender stereotyping

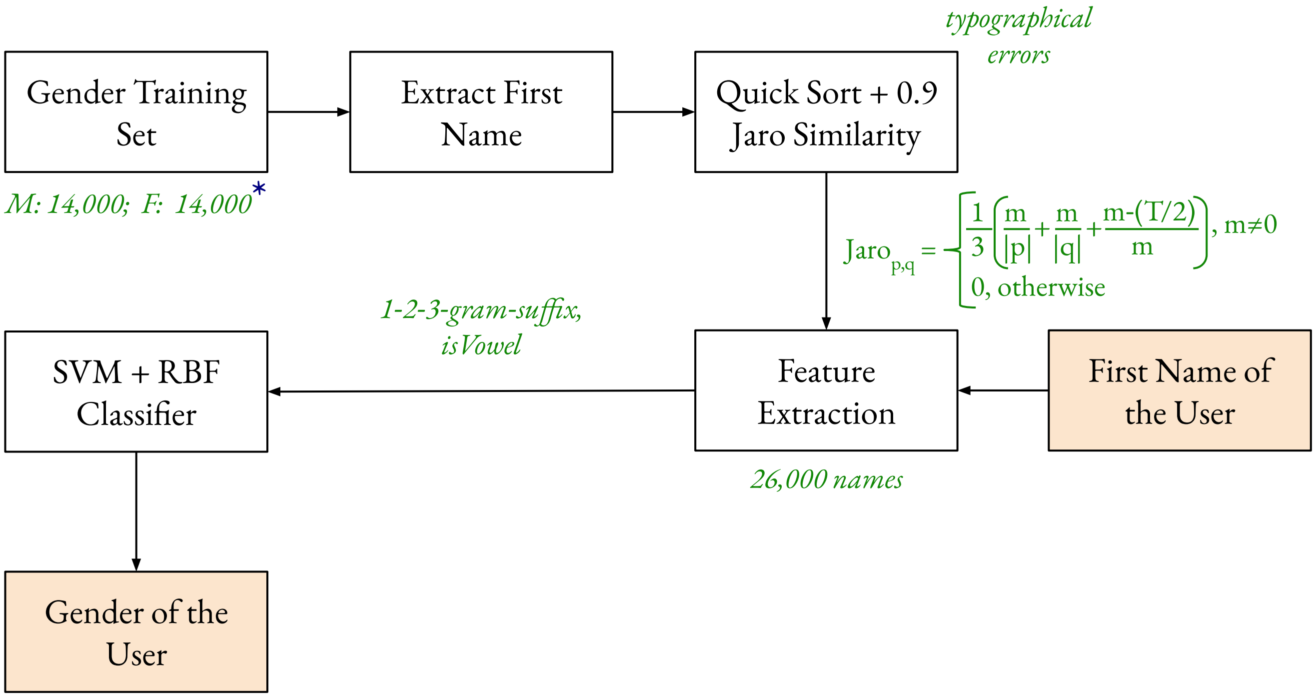


* M: https://gist.github.com/mbejda/7f86ca901fe41bc14a63; F: https://gist.github.com/mbejda/9b93c7545c9dd93060bd

Theories of Social Media Modeling: Gender

- Estimates the social media users' gender to closely mimic the real-world scenario
- Estimate the gender gaps and gender stereotyping

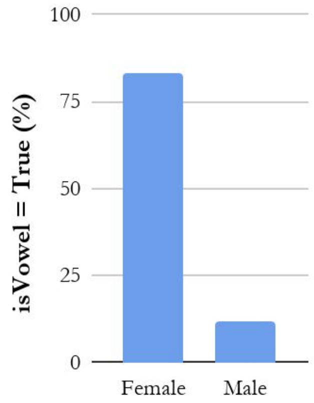


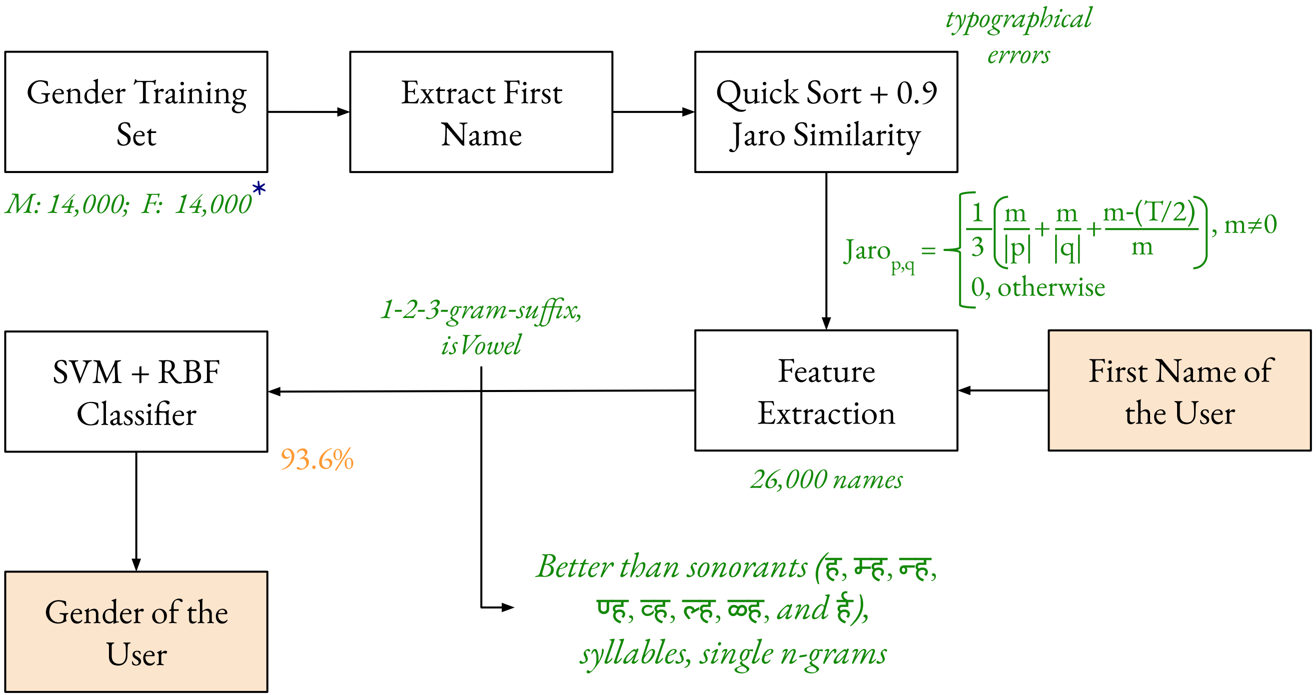


* M: https://gist.github.com/mbejda/7f86ca901fe41bc14a63; F: https://gist.github.com/mbejda/9b93c7545c9dd93060bd

Theories of Social Media Modeling: Gender

- Estimates the social media users' gender to closely mimic the real-world scenario
- Estimate the gender gaps and gender stereotyping



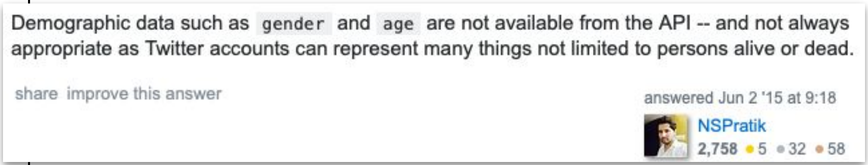


* M: https://gist.github.com/mbejda/7f86ca901fe41bc14a63; F: https://gist.github.com/mbejda/9b93c7545c9dd93060bd

Influential Parameter Mining: Issues and Biases

- Gender and Related Reforms
- Age and Experience
- Religion and Region (geo-tagging)
- Educational backgrounds

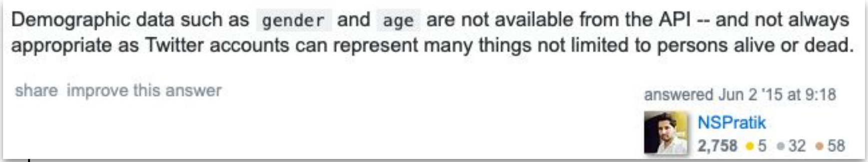
Twitter profile data?



Influential Parameter Mining: Issues and Biases

- Gender and Related Reforms
- Age and Experience
- Religion and Region (geo-tagging)
- Educational backgrounds
- Alliance loyalty (legacy data)
- Impact of political decisions –
  - Farm loan waiver, Minimum support price (agriculture), etc.
- Polling strategies and opposition speeches

Twitter profile data?


Demographic data such as `gender` and `age` are not available from the API -- and not always appropriate as Twitter accounts can represent many things not limited to persons alive or dead.

share improve this answer

answered Jun 2 '15 at 9:18

NSPratik

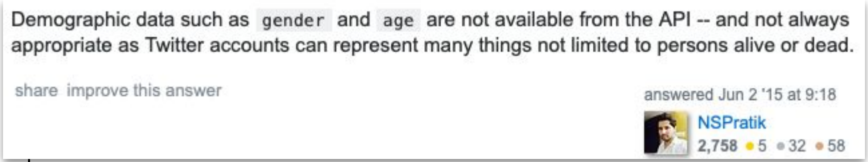
2,758 ●5 ●32 ●58

News articles

Influential Parameter Mining: Issues and Biases

- Gender and Related Reforms
- Age and Experience
- Religion and Region (geo-tagging)
- Educational backgrounds
- Alliance loyalty (legacy data)

Twitter profile data?



Demographic data such as `gender` and `age` are not available from the API -- and not always appropriate as Twitter accounts can represent many things not limited to persons alive or dead.

share improve this answer

answered Jun 2 '15 at 9:18

NSPratik

2,758 ●5 ●32 ●58

- Impact of political decisions –
  - Farm loan waiver, Minimum support price (agriculture), etc.
- Polling strategies and opposition speeches

News articles

- Influence of money, liquor, gifts on elections!
- Financial backgrounds

Cannot be mined for!

Influential Parameter Mining: India

- Pulwama attack+
- Smart cities+
- SC/ST act+
- Ram mandir+
- Swachh Bharath+

Influential Parameter Mining: India

- Pulwama attack^{+}
- Smart cities^{+}
- SC/ST act^{+}
- Ram mandir^{+}
- Swachh Bharath^{+}
- GST^{-}
- Demonetization^{-}
- Black money^{-}
- Ganga clean^{-}

BJP

Influential Parameter Mining: India

- Pulwama attack^{+}
- Smart cities^{+}
- SC/ST act^{+}
- Ram mandir^{+}
- Swachh Bharath^{+}
- GST^{-}
- Demonetization^{-}
- Black money^{-}
- Ganga clean^{-}
- Save India^{+} – Congress+

} BJP

Unification of Modeling Theories: Linear Modeling

- Obtained gender analytics directly influences the volume of the footprints collected (temporally weighted)

Unification of Modeling Theories: Linear Modeling

- Obtained gender analytics directly influences the volume of the footprints collected (temporally weighted)
- Sentiment analysis incorporates temporal weighting and reachability of a tweet*
  - Influential analysis is sentiment analysis of events that affect the voters

* UToSMoV performs a temporal backtracking guided by influence

Unification of Modeling Theories: Linear Modeling

- Obtained gender analytics directly influences the volume of the footprints collected (temporally weighted)
- Sentiment analysis incorporates temporal weighting and reachability of a tweet*
  - Influential analysis is sentiment analysis of events that affect the voters
- UToSMoV score combines volumetric, sentiment, social network, gender, and influential analysis and is normalized per party

$$\mathrm{UToSMoV}(p) = \begin{pmatrix} \beta_{1,m} \\ \beta_{1,f} \end{pmatrix} \cdot \begin{pmatrix} \mathrm{Vol}_m \\ \mathrm{Vol}_f \end{pmatrix}^T + \beta_{2,S} \cdot (\mathrm{Sen}_{pos} - \mathrm{Sen}_{neg}) + \beta_{2,I} \cdot (\mathrm{Inf}_{pos} - \mathrm{Inf}_{neg}) + \beta_3 \cdot (\mathrm{Net}) + \beta_0$$

* UToSMoV performs a temporal backtracking guided by influence

Unification of Modeling Theories: Linear Modeling

- Obtained gender analytics directly influences the volume of the footprints collected (temporally weighted)
- Sentiment analysis incorporates temporal weighting and reachability of a tweet*
  - Influential analysis is sentiment analysis of events that affect the voters
- UToSMoV score combines volumetric, sentiment, social network, gender, and influential analysis and is normalized per party

$$\mathrm{UToSMoV}(p) = \begin{pmatrix} \beta_{1,m} \\ \beta_{1,f} \end{pmatrix} \cdot \begin{pmatrix} \mathrm{Vol}_m \\ \mathrm{Vol}_f \end{pmatrix}^T + \beta_{2,S} \cdot (\mathrm{Sen}_{pos} - \mathrm{Sen}_{neg}) + \beta_{2,I} \cdot (\mathrm{Inf}_{pos} - \mathrm{Inf}_{neg}) + \beta_3 \cdot (\mathrm{Net}) + \beta_0$$

- Constants[1]: $\begin{cases} \beta_{2,S} \leq \beta_3 \leq \beta_{2,I} \cong (\beta_{1,m} + \beta_{1,f}), \text{ if } \beta_{1,m} \text{ and } \beta_{1,f} \text{ are known} \\ \beta_{2,S} \leq \beta_3 \leq \beta_{2,I} \text{ and } (\beta_{1,m} = \beta_{1,f} = 1), \text{ otherwise} \end{cases}$

* UToSMoV performs a temporal backtracking guided by influence

Unification of Modeling Theories: Linear Modeling

- Obtained gender analytics directly influences the volume of the footprints collected (temporally weighted)
- Sentiment analysis incorporates temporal weighting and reachability of a tweet*
  - Influential analysis is sentiment analysis of events that affect the voters
- UToSMoV score combines volumetric, sentiment, social network, gender, and influential analysis and is normalized per party

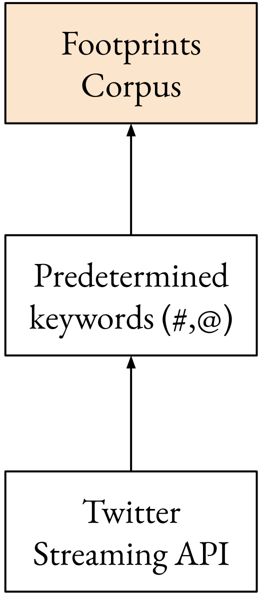
$$\text{UToSMoV}(p) = \begin{pmatrix} \beta_{1,m} \\ \beta_{1,f} \end{pmatrix} \cdot \begin{pmatrix} \text{Vol}_m \\ \text{Vol}_f \end{pmatrix}^T + \underbrace{\beta_{2,S}}_{\boxed{0.7}} \cdot (\text{Sen}_{pos} - \text{Sen}_{neg}) + \underbrace{\beta_{2,I}}_{\boxed{1.0}} \cdot (\text{Inf}_{pos} - \text{Inf}_{neg}) + \underbrace{\beta_3}_{\boxed{0.8}} \cdot (\text{Net}) + \beta_0$$

- Constants[1]: $\begin{cases} \beta_{2,S} \leq \beta_3 \leq \beta_{2,I} \cong (\beta_{1,m} + \beta_{1,f}), \text{ if } \beta_{1,m} \text{ and } \beta_{1,f} \text{ are known} \\ \beta_{2,S} \leq \beta_3 \leq \beta_{2,I} \text{ and } (\beta_{1,m} = \beta_{1,f} = 1), \text{ otherwise} \end{cases}$

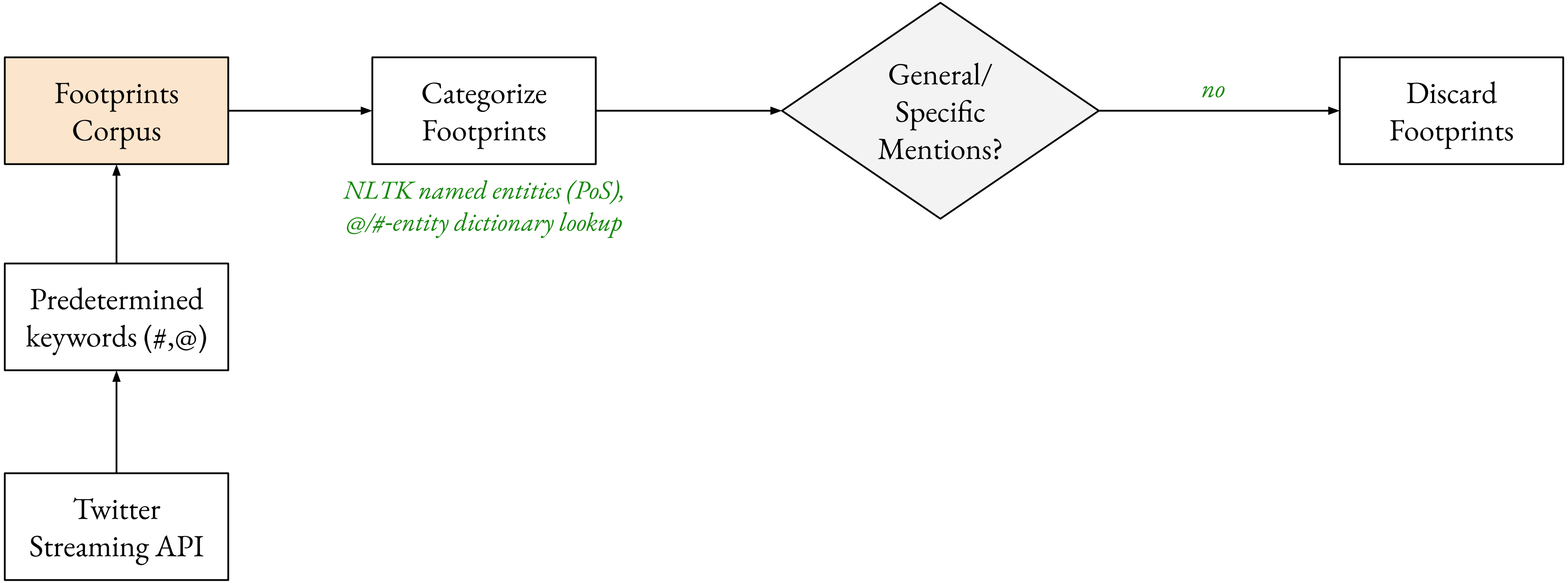
* $0 \leq \beta_{1,m}; \beta_{1,f}; \beta_{2,S}; \beta_{2,I}; \beta_3 \leq 1$

* UToSMoV performs a temporal backtracking guided by influence

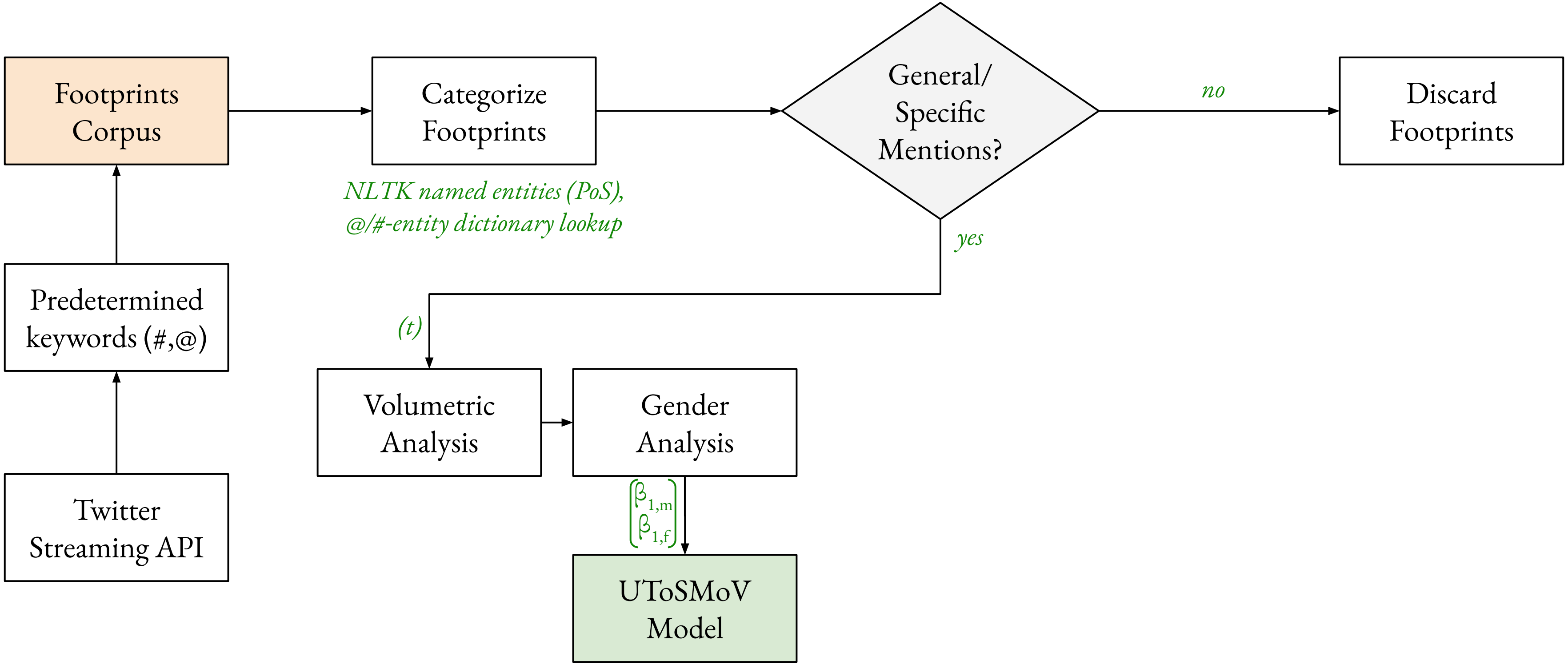
Unification of Modeling Theories: UToSMoV



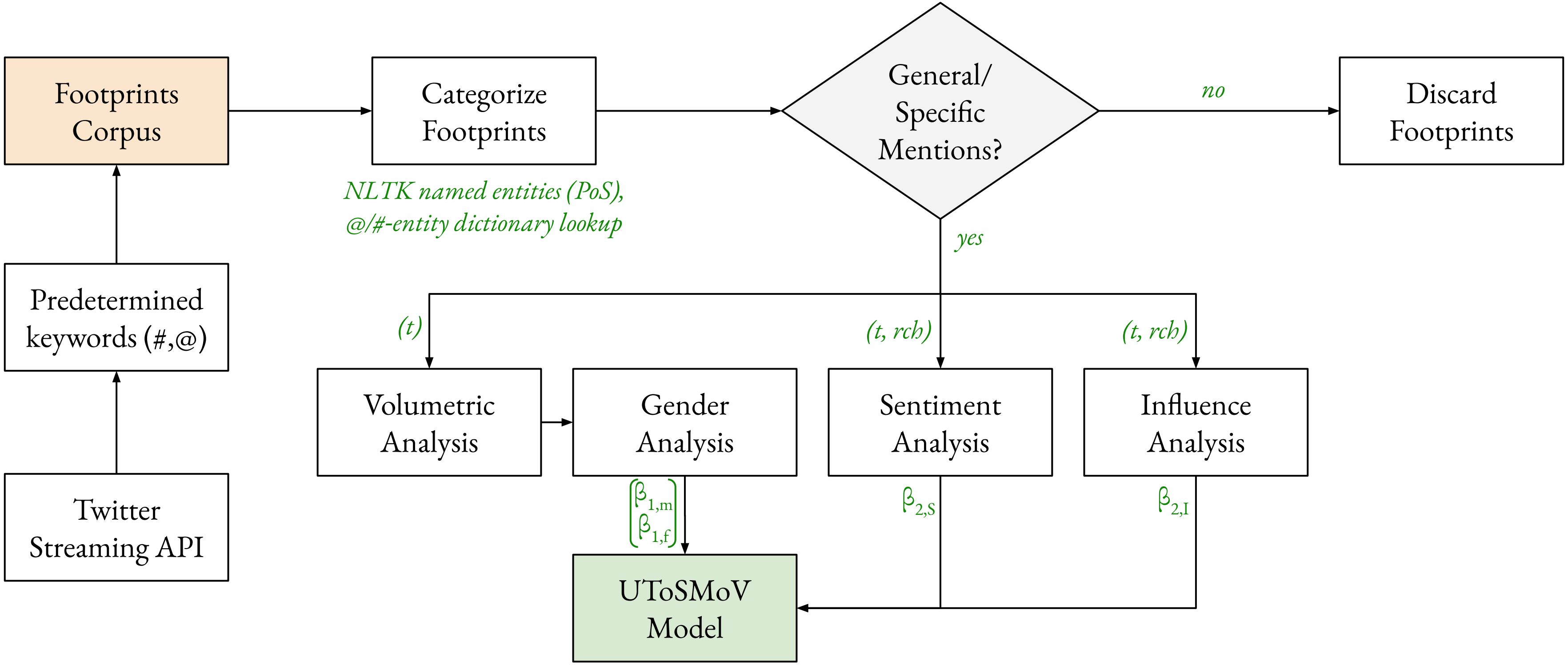
Unification of Modeling Theories: UToSMoV



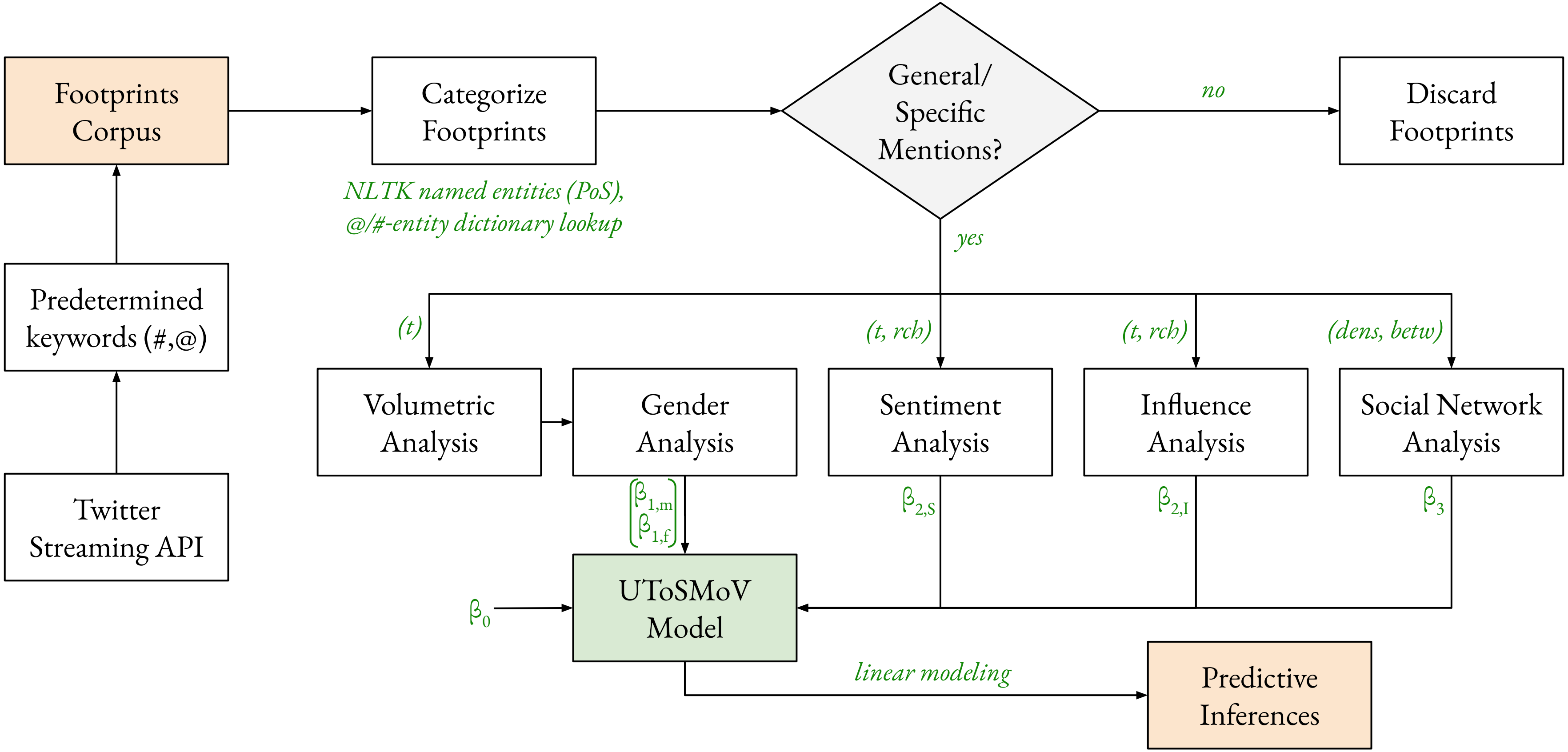
Unification of Modeling Theories: UToSMoV



Unification of Modeling Theories: UToSMoV



Unification of Modeling Theories: UToSMoV



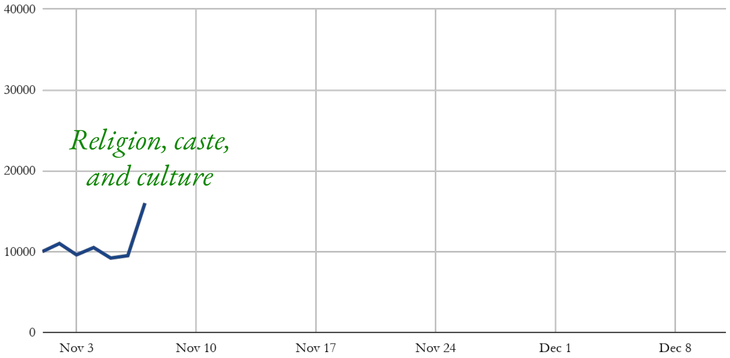
Twitter Corpus Statistics

State	Political Party	Tweets Collected		Number of Handles*		Collection Time (days)
Telangana	BJP	23,642	241,489	4	19	~35
	Congress+	55,751		7		~35
	TRS	162,096		8		~35
Chhattisgarh	BJP	4,317	7,674	16	27	~25
	Congress+	3,357		11		~25
Rajasthan	BJP	101,040	168,228	28	55	~30
	Congress+	67,188		27		~30
Madhya Pradesh	BJP	118,197	178,313	24	39	~30
	Congress+	60,116		15		~30

* Other generic handles were also used to collect the data and were then classified into a particular party based on the user mentions

Twitter Corpus Statistics

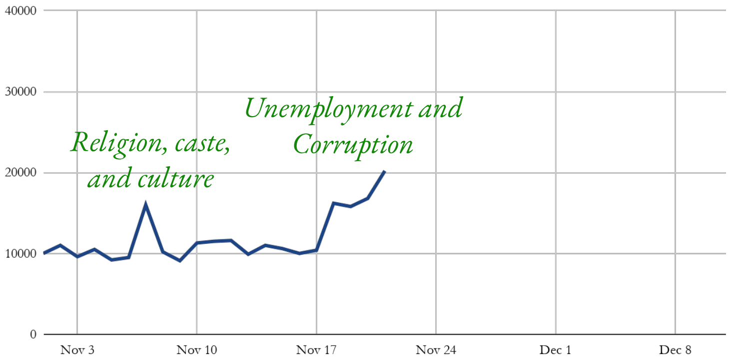
State	Political Party	Tweets Collected		Number of Handles*		Collection Time (days)
Telangana	BJP	23,642	241,489	4	19	~35
	Congress+	55,751		7		~35
	TRS	162,096		8		~35
Chhattisgarh	BJP	4,317	7,674	16	27	~25
	Congress+	3,357		11		~25
Rajasthan	BJP	101,040	168,228	28	55	~30
	Congress+	67,188		27		~30
Madhya Pradesh	BJP	118,197	178,313	24	39	~30
	Congress+	60,116		15		~30



* Other generic handles were also used to collect the data and were then classified into a particular party based on the user mentions

Twitter Corpus Statistics

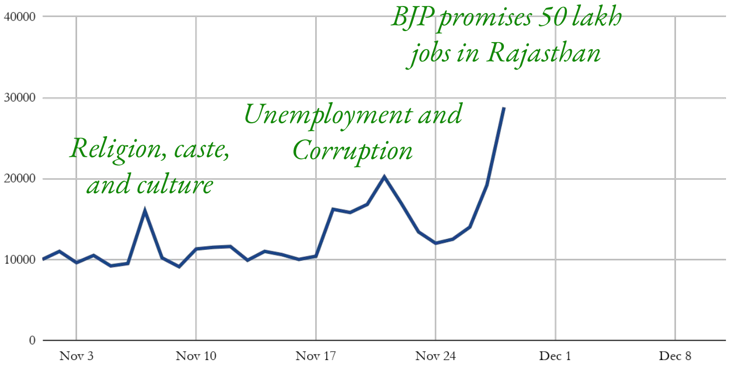
State	Political Party	Tweets Collected		Number of Handles*		Collection Time (days)
Telangana	BJP	23,642	241,489	4	19	~35
	Congress+	55,751		7		~35
	TRS	162,096		8		~35
Chhattisgarh	BJP	4,317	7,674	16	27	~25
	Congress+	3,357		11		~25
Rajasthan	BJP	101,040	168,228	28	55	~30
	Congress+	67,188		27		~30
Madhya Pradesh	BJP	118,197	178,313	24	39	~30
	Congress+	60,116		15		~30



* Other generic handles were also used to collect the data and were then classified into a particular party based on the user mentions

Twitter Corpus Statistics

State	Political Party	Tweets Collected		Number of Handles*		Collection Time (days)
Telangana	BJP	23,642	241,489	4	19	~35
	Congress+	55,751		7		~35
	TRS	162,096		8		~35
Chhattisgarh	BJP	4,317	7,674	16	27	~25
	Congress+	3,357		11		~25
Rajasthan	BJP	101,040	168,228	28	55	~30
	Congress+	67,188		27		~30
Madhya Pradesh	BJP	118,197	178,313	24	39	~30
	Congress+	60,116		15		~30



* Other generic handles were also used to collect the data and were then classified into a particular party based on the user mentions

Twitter Corpus Statistics

State	Political Party	Tweets Collected		Number of Handles*		Collection Time (days)
Telangana	BJP	23,642	241,489	4	19	~35
	Congress+	55,751		7		~35
	TRS	162,096		8		~35
Chhattisgarh	BJP	4,317	7,674	16	27	~25
	Congress+	3,357		11		~25
Rajasthan	BJP	101,040	168,228	28	55	~30
	Congress+	67,188		27		~30
Madhya Pradesh	BJP	118,197	178,313	24	39	~30
	Congress+	60,116		15		~30

* Other generic handles were also used to collect the data and were then classified into a particular party based on the user mentions

Twitter Corpus Statistics

State	Political Party	Tweets Collected		Number of Handles*		Collection Time (days)
Telangana	BJP	23,642	241,489	4	19	~35
	Congress+	55,751		7		~35
	TRS	162,096		8		~35
Chhattisgarh	BJP	4,317	7,674	16	27	~25
	Congress+	3,357		11		~25
Rajasthan	BJP	101,040	168,228	28	55	~30
	Congress+	67,188		27		~30
Madhya Pradesh	BJP	118,197	178,313	24	39	~30
	Congress+	60,116		15		~30

- Top mentioned leaders:
  - @narendramodi
  - @RahulGandhi
  - @AmitShah
  - @myogiadityanath
  - @yadavakhilesh

* Other generic handles were also used to collect the data and were then classified into a particular party based on the user mentions

Twitter Corpus Statistics

State	Political Party	Tweets Collected		Number of Handles*		Collection Time (days)
Telangana	BJP	23,642	241,489	4	19	~35
	Congress+	55,751		7		~35
	TRS	162,096		8		~35
Chhattisgarh	BJP	4,317	7,674	16	27	~25
	Congress+	3,357		11		~25
Rajasthan	BJP	101,040	168,228	28	55	~30
	Congress+	67,188		27		~30
Madhya Pradesh	BJP	118,197	178,313	24	39	~30
	Congress+	60,116		15		~30

- Top mentioned leaders:
  - @narendramodi
  - @RahulGandhi
  - @AmitShah
  - @myogiadityanath
  - @yadavakhilesh
- Top conversations:
  - Rural economy
  - Religion and Caste
  - Vote tampering
  - Dynastic politics
  - Corruption

* Other generic handles were also used to collect the data and were then classified into a particular party based on the user mentions

UToSMoV: Data preprocessing

Repeated words (> 4)

Punctuations (>5)

Retweets

~~Word smoothing~~

Internet slang*

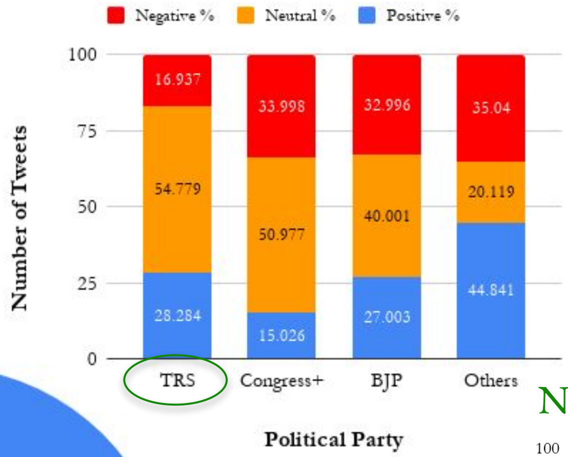
#	User Twitter ID	User Screen Name	User Gender	User Follow. Count	User Friends Count	User Listed Count	User Location	Tweet Created At	Tweet Lang.
4	1052824183866982400	Vinay Bhaskar	male	6	79	2	Andhra Pradesh India	Fri Nov 09 04:27:26 +0000 2018	te

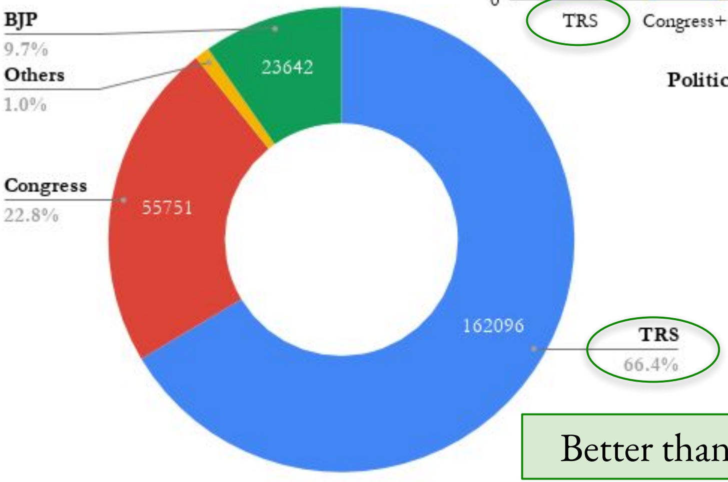
Tweet Hashtags	Tweet User Mentions	Tweet Retweet Count	Tweet Favorite Count	Tweet Quote Count	Tweet Reply Count	Tweet Text	Tweet Senti.	Tweet Alliance
#SaveTelangana #SaveDemocracy	@PTelangana @KTRTRS @RaoKavitha @trsharish @sushilrTOI	20	200	5	20	RT @PTelangana: KCR పార్టీ అరాచకాలు...@KTRTRS @RaoKavitha @trshar...	negative	TRS

* Abbreviation library can be found at: https://www.netlingo.com/acronyms.php

UToSMoV on Telangana: Results and Analysis

Sentiment Analysis ↦

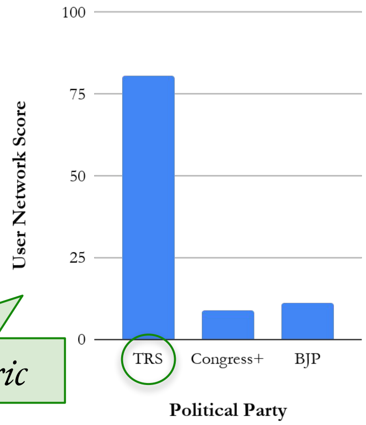


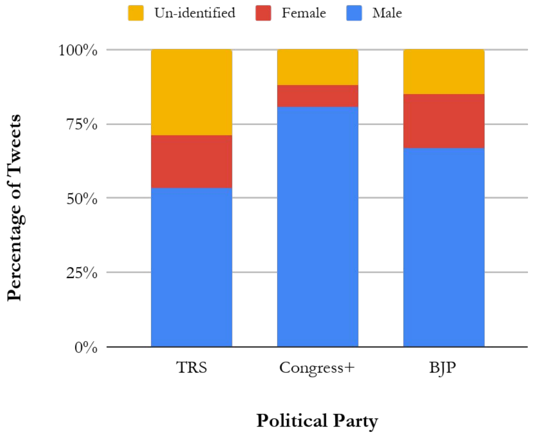


Volumetric Analysis ↥

Better than *Volumetric*

Network Analysis ↧





Gender Analysis ↥

Political engagement
Cognition *vs.* group thinking
Mass level voting behaviour

Male	: 50.3%
Female	: 48.0%
Others	: 01.7%

http://www.indiavotes.com/state/summary/61

UToSMoV on Telangana: Influential Parameters

- Kaleshwaram
- AP reorganisation act
- Mission Kakatiya
- Mission Bhagiratha
- Hyderabad metro rail
- Reservation bill act

TRS

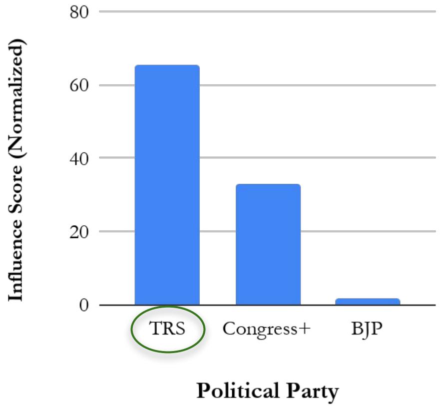
- Indiramma Illu
- One-lakh obs
- 30 Days 30 Questions

Congress+

- Cows distribution
- Renaming cities

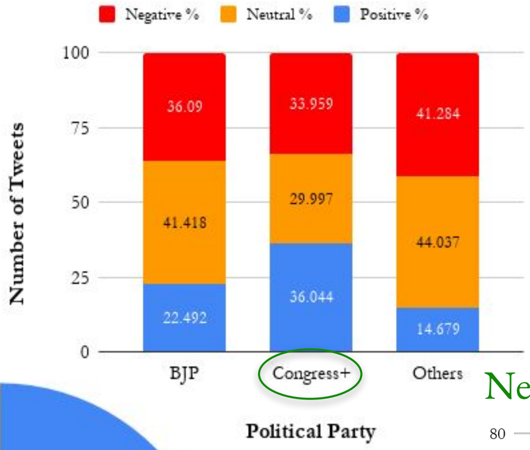
BJP

@KTRTRS,
@asadowaisi,
@UttamTPCC,
@drlaxmanbjp

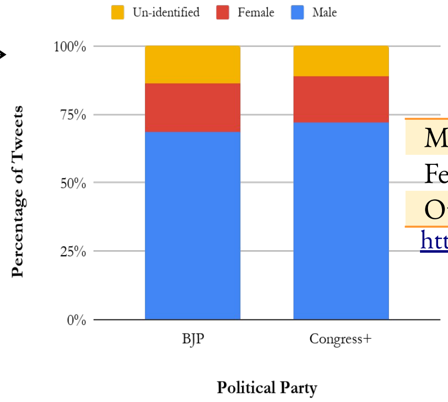


UToSMoV on Chhattisgarh: Results and Analysis

Sentiment Analysis ↦

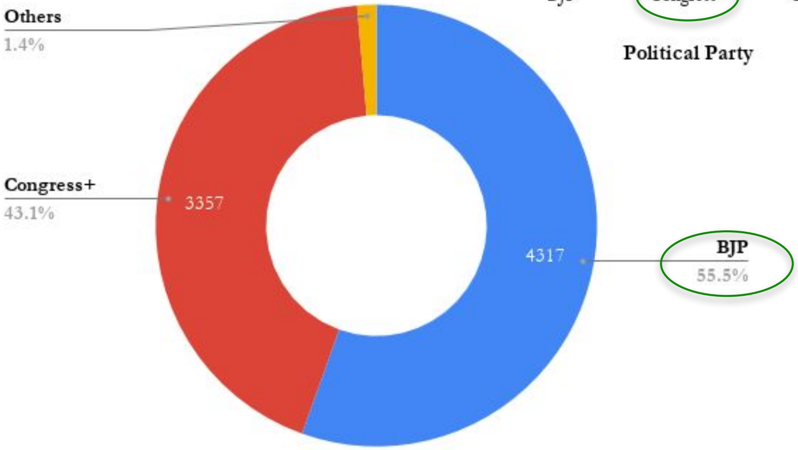


Gender Analysis ↦



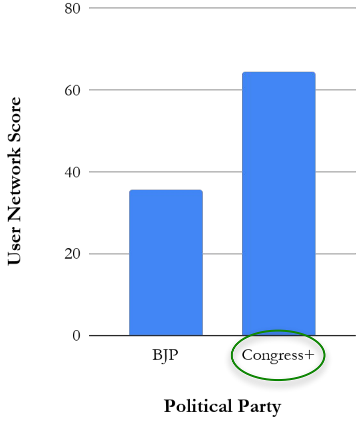
Male	: 50.5%
Female	: 49.5%
Others	: 00.0%

http://www.indiavotes.com/state/summary/54

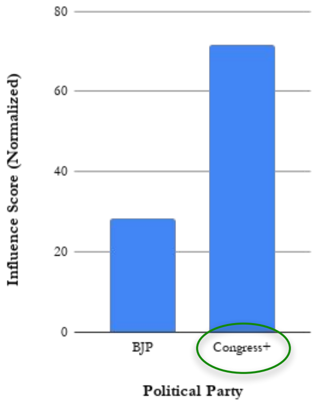


Volumetric Analysis ↥

Network Analysis ↧



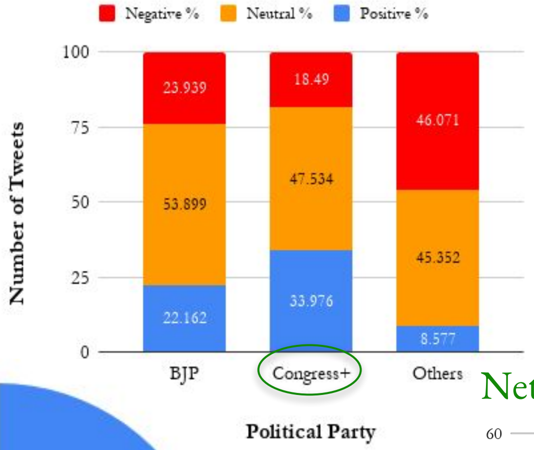
Influential Analysis ↧



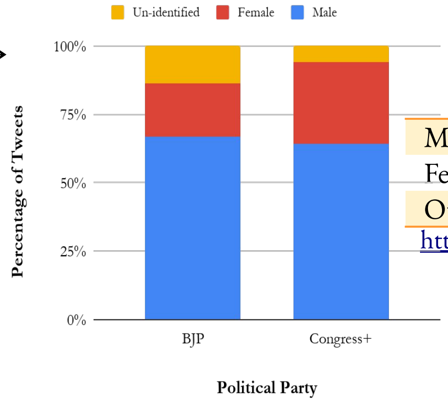
@drramansingh,
@Bhupesh_Baghel,
@ajitjogi_cg

UToSMoV on Rajasthan: Results and Analysis

Sentiment Analysis ↦

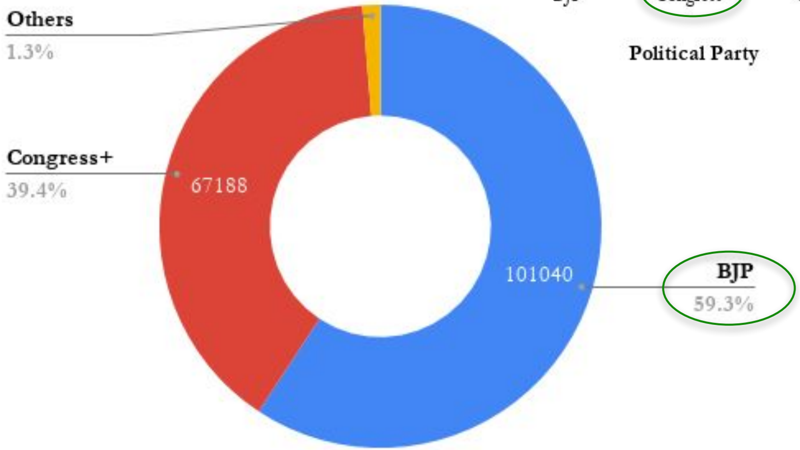


Gender Analysis ↦



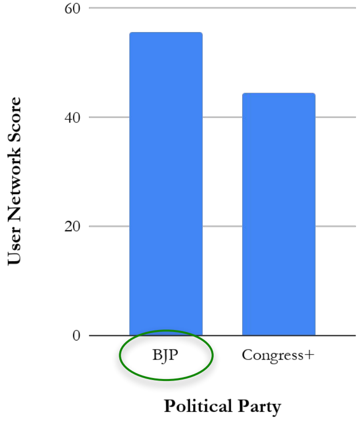
Male	: 52.7%
Female	: 47.3%
Others	: 00.0%

http://www.indiavotes.com/state/summary/14

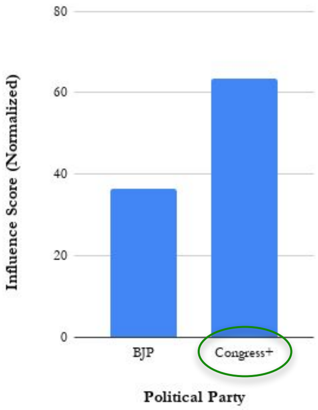


Volumetric Analysis ↥

Network Analysis ↧



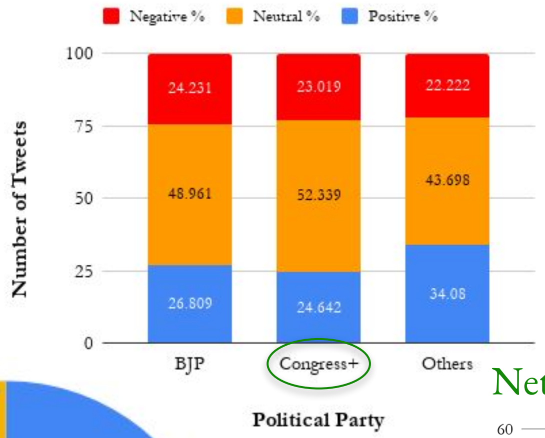
Influential Analysis ↧



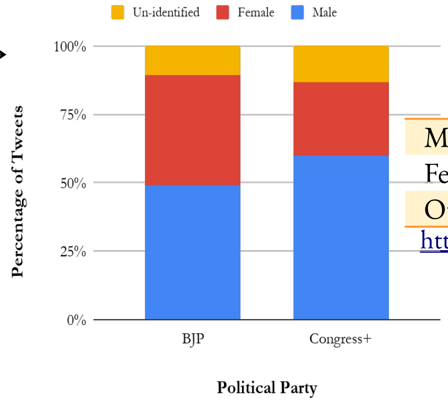
@VasundharaBJP,
@SachinPilot,
@ashokgehlot51,
@MPMadanSaini

UToSMoV on Madhya Pradesh: Results and Analysis

Sentiment Analysis ↦

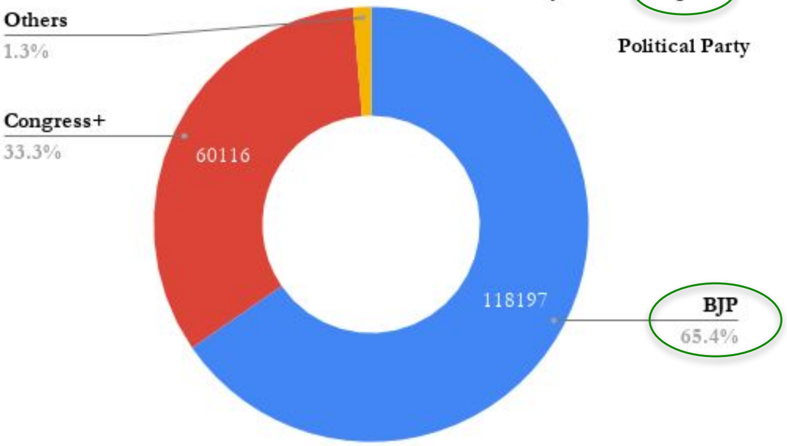


Gender Analysis ↦



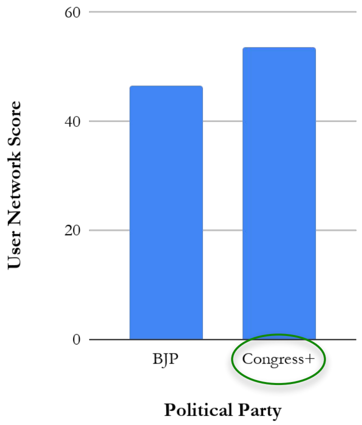
Male	: 52.7%
Female	: 47.3%
Others	: 00.0%

http://www.indiavotes.com/state/summary/59

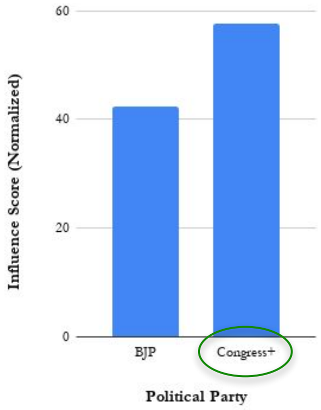


Volumetric Analysis ↥

Network Analysis ↧

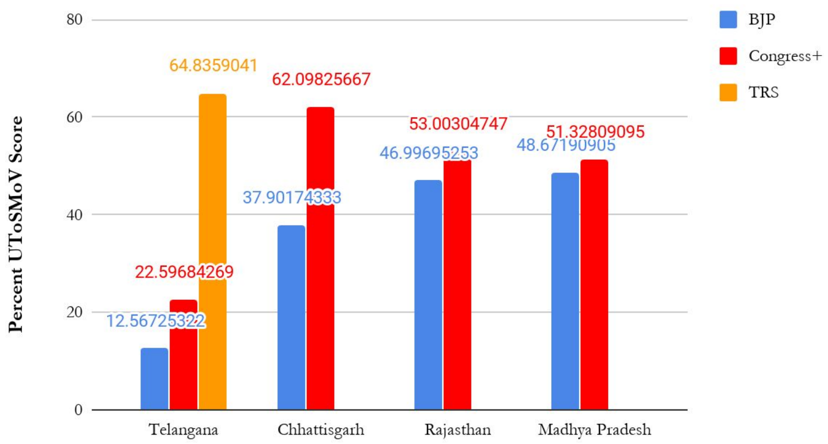


Influential Analysis ↧



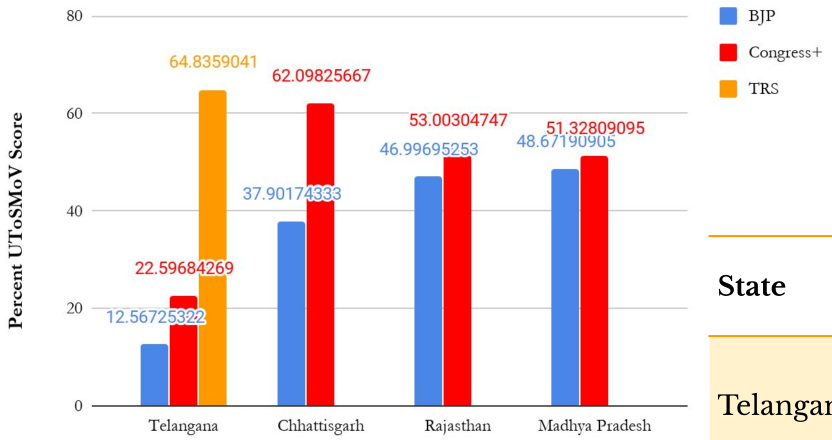
@ChouhanShivraj,
@JM_Scindia,
@OfficeOfKNath,
@ChitnisArchana

Unification of Various Theories: Predictions



- Telangana: TRS
- Chhattisgarh: Congress
- Rajasthan: Congress
- Madhya Pradesh: Congress

Unification of Various Theories: Quantified Selfie



- Telangana: TRS
- Chhattisgarh: Congress
- Rajasthan: Congress
- Madhya Pradesh: Congress

State	Political Party	Times Now CNX Exit Poll	C-Voter Exit Poll	UToSMoV	Actual Result
Telangana	BJP	05.88%	04.20%	12.57%	00.84%
	Congress+	31.09%	37.82%	22.60%	17.64%
	TRS	55.46%	48.74%	64.84%	73.95%
Chhattisgarh	BJP	51.11%	41.60%	37.90%	24.44%
	Congress+	38.89%	42.20%	62.10%	75.55%
Rajasthan	BJP	42.71%	39.70%	47.00%	39.69%
	Congress+	52.76%	47.90%	53.00%	50.25%
Madhya Pradesh	BJP	54.78%	41.50%	48.67%	48.26%
	Congress+	38.70%	42.30%	51.33%	49.56%

Conclusions and Future Work

- Unified framework that models volumetric, sentiment, social network, gender, and influence outperforms the baseline predictions
- Temporal backtracking guided by influence accounts for a change in the user's opinion due to the political party's influence
- Fact checking and fake-news detection modules are to be incorporated to enable more accurate predictions
- An effective strategy for bot and sockpuppet identification must be developed
- Develop a parameter self-adaptive model to learn the unification parameters
- Post-election alliances to be found using legacy data via monte-carlo simulations
- Polling strategies such as opposition speeches are to be analyzed

References

[1] Mukku, S.S., and Mamidi, R. "ACTSA: Annotated corpus for telugu sentiment analysis." Proceedings of the First Workshop on Building Linguistically Generalizable NLP Systems. 2017.

[2] Akhtar, M.S., et al. "A hybrid deep learning architecture for sentiment analysis." Proceedings of COLING 2016, the 26th International Conference on Computational Linguistics: Technical Papers. 2016.

[3] Ameya, P., et al. "Towards sub-word level compositions for sentiment analysis of hindi-english code mixed text." arXiv preprint arXiv:1611.00472 (2016).

[4] Jaidka, K., et al. "Predicting elections from social media: a three-country, three-method comparative study." Asian Journal of Communication (2018): 1-21.

[5] Cuzzocrea, A., et al. "Edge betweenness centrality: A novel algorithm for QoS-based topology control over wireless sensor networks." Journal of Network and Computer Applications (2012).