How Constituents Lobby Members of Congress on Twitter

Andrew Roback and Libby Hemphill
Illinois Institute of Technology

Twitter and politics: a brief history

- Twitter: a social networking site that allows posts of 140 characters or less
- Including usernames (e.g. @SenatorKirk) allows users to directly target MOCs
- **2009:** 158 MOCs with Twitter accounts
- **2012:** <u>every</u> newly elected MOC had a Twitter account

Glassman, Straus, & Shogan 2010; Hemphill & Roback, forthcoming

Twitter as compared with traditional communication channels

- More public than email or letter writing
- Very short, necessitating directness from users
- #Hashtags quickly identify issues and enable the watching of discourse on specific issues

Related work in CMC

- How citizens talk to each other on SNS
  - Mascaro, Black, & Goggins, 2012; Morgan, Lampe, & Shafiq, 2013; Munson & Resnick, 2011
- How MOCs use SNS to talk to constituents
  - Golbeck et al., 2010; Hemphill, Otterbacher, & Shapiro, 2013
- How constituents lobby MOCs
  - Roback & Hemphill, 2013; Roback and Hemphill, forthcoming

Problem

- Citizens talk a lot about politics online
- MOCs are present on Twitter, but don’t dialogue with citizens very much
- How do citizens talk to MOCs?
- What approaches get the best (or any) feedback?

Problem

In a time when social media is used for popular empowerment globally, yet our own unpopular Congress is on track to become the least productive in modern history[1], **how does an engaged electorate on Twitter lobby for legislation** that addresses important national and social issues?

[1] http://www.washingtonpost.com/blogs/the-fix/wp/2013/08/02/judging-the-unproductivity-of-the-113th-congress/

Our approach

- Rhetorically and linguistically situate Twitter as a method of discourse
- Collect tweets and group by common approaches
- Code a subset and use a machine learning algorithm to code the larger set

Rhetorically situating lobbying on Twitter

- Aristotle defines three distinct species of rhetoric
  - Judicial: matters of innocence v. guilt
  - Epideictic: matters of praise v. blame
  - Deliberative: determining an advantageous course of action

- Arguments arise over which course of action is more advantageous

Rhetorically situating lobbying on Twitter

- A special type of argument, or *koina*, concerned with the concept of more and less
- A *pistis*, or rhetorical proof, that can be used with this type of argument:

“what all people prefer is preferable to what all do not. And what more rather than fewer prefer is preferable; for *good* was what all desire so *greater* is what more people desire”

(Aristotle, I 7:28)

Rhetorically situating lobbying on Twitter

- Aristotle defines three distinct species of rhetoric
  - Judicial: matters of innocence v. guilt
  - Epideictic: matters of praise v. blame
  - Deliberative: determining the *degree* of goodness in a course of action (good v. better)

Rhetorically situating lobbying on Twitter

- Power in numbers as a fundamental assumption
- Assumes MOCs will change position to reflect the *vox populi*
- Does not fully explain the varied and sophisticated strategies constituents use

Linguistically situating lobbying on Twitter

- Speech act theory
  - Speakers have intent and try to achieve some effect
  - Speakers make utterances “[not] merely to exercise their vocal cords” (Bach & Harnish, 1979) but to achieve some effect.
  - We translate this concept to Tweets

Linguistically situating lobbying on Twitter

- Searle's categories of speech acts
  - **directives**, which attempt to get the listener to do something;
  - **commissives**, which commit the speaker to a course of action;
  - **representatives**, which serve to report on the state of the world;
  - **expressives**, which express a speaker's emotional state; and
  - **declarations**, which change the state of a person or object (e.g. saying “I resign” actually changes your status as an employee)

(Searle, 1976)

Linguistically situating lobbying on Twitter

- Searle’s categories of speech acts
  - **directives**, which attempt to get the listener to do something;
  - **commissives**, which commit the speaker to a course of action;
  - **representatives**, which serve to report on the state of the world;
  - **expressives**, which express a speaker's emotional state; and
  - ~~**declarations**, which change the state of a person or object (e.g. saying “I resign” actually changes your status as an employee)~~
  - **questions**, which attempt to solicit information from the hearer

(Searle, 1976; Parker & Riley, 2011)

Rhetoric and linguistics

“Building on Aristotle's conceptions of rhetoric, we can say that Twitter is a space where conceptions of mass desire and vox populi *are important, but where individuality and uniqueness of appeal also count for something.”*

Rhetoric and linguistics

“Building on Aristotle's conceptions of rhetoric, we can say that Twitter is a space where conceptions of mass desire and vox populi *are important, but where individuality and uniqueness of appeal also count for something.”*

- Not surprising, as we demonstrate variety of appeal in everyday interactions

Methods

- Collected **76,454 tweets** from **43,079 users** directed at **566 accounts** owned by MOCs
- **Less retweets**, that left **34,056 tweets**

Issue	Hashtags	Tweets	Users
Immigration reform	#immigration	4845	3083
	#dreamact	2591	1838
	#dreamers	3175	2495
Federal budget and the sequester	#budget	13,249	8767
	#fiscalcliff	978	674
	#sequestration	914	647
Gun control	#guncontrol	1743	733
	#2ndamendment	1443	1014
	#nra	1819	747
Internet freedom	#sopa	36,985	21,265
	#pipa	25,009	15,633
	#cispa	5498	3712
Total		76,454*	43,079*

* Some tweets contained multiple hashtags, and some users posted more than one tweet. These numbers represent unique tweets and unique users.

Methods

- Developed a set of 16 common lobbying strategies based rhetorical approach and speech act theory

Methods

1. I’d have to vote against you…
2. Directly oppose/support
3. FYI
4. Please oppose/support
5. General directive
6. Thank you for opposing/supporting
7. Disappointed
8. I want a response from you
9. Loaded policy question
10. Rhetorical question
11. What is your position?
12. Promotional
13. Campaign ad accusation
14. I’m your constituent and I oppose
15. Analogy
16. Other

Methods

1. I'd have to vote against you…	**Commissive**
2. Directly oppose/support 3. FYI 4. Please oppose/support 5. General directive	**Directive**
6. Thank you for opposing/supporting 7. Disappointed 8. I want a response from you	**Expressive**
9. Loaded policy question 10. Rhetorical question 11. What is your position?	**Question**
12. Promotional 13. Campaign ad accusation 14. I'm your constituent and I oppose 15. Analogy	**Representative**
16. Other	**N/A**

Methods - Coding by hand

- Coded a sample set and achieved substantial agreement (k=0.73)
- Independently coded a random set of 669 tweets

(Landis & Koch, 1977)

Methods - Coding algorithmically

- Used the human coded tweets as a training set
- Remaining 33,387 are test set
- Trained a variety of classifiers, including naive Bayes, J48, Decision Table, and Bayes net

Methods - Coding algorithmically

- naive Bayes uses the bag of words approach
    - Cannot handle multi-word strings

@SenJohnMcCain Please, put #GunControl in your agenda. No more weak NRA laws. We can stop future massacres. #Aurorashootings #Colorado

Methods - Coding algorithmically

@SenJohnMcCain Please, put #GunControl in your agenda. No more weak NRA laws. We can stop future massacres. #Aurorashootings #Colorado

transform to a list text attributes

@SenJohnMcCain	your	stop
Please	agenda	future
put	weak	massacres
#GunControl	NRA	#Aurorashootings
in	laws	#Colorado
	We	

Methods - Coding algorithmically

- Validation
  - Assessing accuracy on test data is misleading, hence the need for validation
  - We use stratified, 10-fold cross validation

Results

- Classifier performed poorly on the 16 class data configuration
  - accuracy = 46%
  - k = 0.39 (fair agreement)

What about a speech acts classifier?

- Reconfigured to a five class data set corresponding to five speech act types
  - N/A, or no discernable speech act excluded (n = 54)
  - speech acts training set (n = 615)

What about a speech acts classifier?

- Reconfigured to a five class data set corresponding to five speech act types
  - N/A, or no discernable speech act excluded (n = 54)
  - speech acts training set (n = 615)
- Can we further transform data set to improve accuracy?
  - One-versus-all technique
    - Transforms a multi-class classification task into n-binary classifications tasks

Results

- Speech act classifier performed better
  - accuracy = 62%
  - k = 0.47 (moderate agreement)
  - probably due in no small part to data transformations

Speech act type	Human Coded Training Set		Algorithmically Coded Set	
	N	%	N	%
Commissive	18	3%	534	2%
Directive	241	39%	22465	66%
Expressive	105	17%	2152	6%
Question	95	16%	6359	19%
Representative	156	25%	2546	7%

Discussion

- Indeed, people do use sophisticated appeals when lobbying MOCs
    - Directive prevailed as dominant, but other speech acts well represented
- Our speech acts classifier is still minimally useful, but a good starting point
    - What hampered accuracy in our classifiers?
    - What can we do about it in future experiments?

Discussion: What hampered accuracy?

Three major issues:

1. Number of categories
2. Length of documents
3. Context of speech

Discussion:
What hampered accuracy?

1. Number of categories
   - As mutually exclusive categories increase:
     - Error probability increases
     - Highly predictive text attributes are dampened
   - Less classes means fewer chances to make an error
     - Probably accounts for some of the improvement from rhetorical appeals (classes = 16) to speech acts (classes = 5)

Discussion:
What hampered accuracy?

2. Length of documents

- Tweets are short, hence less words to associate with each class
    - When a highly predictive attribute is associated with two classes, the classifier gets confused

Discussion: What hampered accuracy?

2. Length of documents

- Finding predictive attributes
  - Correlation-based feature selection (Hall, 1999) with ten-fold cross validation

Percent of folds in which attribute was highly predictive	Attribute
100	#fiscalcliff
100	#guncontrol
100	for
100	standing
100	you
100	Thank / thank
100	where
100	How
90	http

topic word

Discussion:
What hampered accuracy?

@SenJohnMcCain Please, put #GunControl in your agenda. No more weak NRA laws. We can stop future massacres. #Aurorashootings #Colorado

transform to a list text attributes

```
@SenJohnMcCain   your      stop
Please           agenda    future
put              weak      massacres
#GunControl      NRA       #Aurorashootings
in               laws      #Colorado
                 We
```

Discussion:
What hampered accuracy?

@SenJohnMcCain	your	**stop**
Please	agenda	future
put	weak	massacres
#GunControl	NRA	#Aurorashootings
in	laws	#Colorado
	We	

Directive

4. Please oppose/support

Discussion:
What hampered accuracy?

@SenJohnMcCain
Please
put
#GunControl
in
your
agenda
weak
NRA
laws
We
stop
future
massacres
#Aurorashootings
#Colorado

Um...

Discussion: What hampered accuracy?

2. Length of documents

- Finding predictive attributes
  - Correlation-based feature selection (Hall, 1999) with ten-fold cross validation

Percent of folds in which attribute was highly predictive	Attribute
100	#fiscalcliff
100	#guncontrol
100	for
100	standing
100	you
100	Thank / thank
100	where
100	How
90	http

topic word

expressive

question

Discussion:
What hampered accuracy?

3. Context of speech
   - Contextual or non-literal utterances
     - sarcasm
     - analogies
   - Polarity reversal and negation
     - “I sure **don’t like** Senator Durbin’s #immigration” stance”
     - **“I sure** don’t like Senator **Durbin’s** ~~#immigration~~” stance”

Discussion:
What hampered accuracy?

3. Context of speech

- Contextual or non-literal utterances
  - sarcasm
  - analogies
- Polarity reversal and negation
  - "I sure **don't like** Senator Durbin's #immigration" stance"
    7. Disappointed
  - "**I sure** don't like Senator **Durbin's** ~~**#immigration**~~" stance"
    Um...

Future work

- Accounting for number of categories
  - Replace mutually exclusive classes with classes that allow instances in more than one category
  - Identify primary and secondary rhetorical approaches

Future work

- Accounting for short document length / predictive attribute problem
  - Combine training instances into larger documents to increase dictionary for each class
  - Balance training set by issue type, hashtag, and class occurrence
    - We did this for class, but only by removing instances
  - Separate informational (#GunControl) from semantic content

Future work

- Accounting for context
  - More complex, but not impossible. Headway has been made in the field of sentiment analysis

Future work

- Moving forward
  - Investigate the responses that MOCs give based on various rhetorical techniques and speech act types
  - Important to understand what type of approach elicits a favorable (or any) response

How Constituents Lobby Members of Congress on Twitter

Andrew Roback and Libby Hemphill
Illinois Institute of Technology

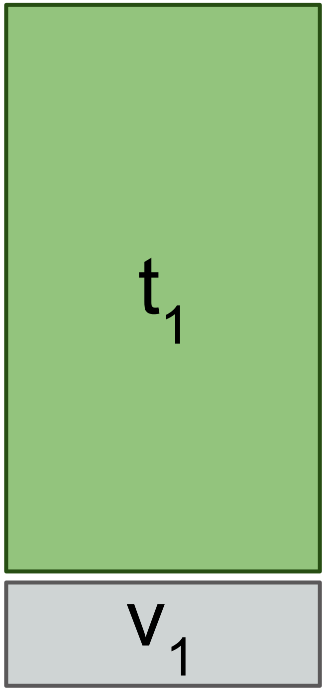
Hand Coding Results

Speech Act Type	N (%)
Commissive	18 (3%)
Directive	241 (36%)
Expressive	105 (16%)
Questions	95 (14%)
Representative	156 (23%)
N/A	54 (8%)

Code	N (%)
I'd have to vote against you…	18 (3%)
Directly oppose/support	117 (17%)
FYI	76 (11%)
Please oppose/support	30 (4%)
General directive	18 (3%)
Thank you for opposing/supporting	71 (11%)
Disappointed	20 (3%)
I want a response from you	14 (2%)
Loaded policy question	42 (6%)
Rhetorical question	33 (5%)
What is your position?	20 (3%)
Promotional	85 (13%)
Campaign ad accusation	55 (8%)
I'm your constituent and I oppose	11 (2%)
Analogy	5 (1%)
Other	54 (8%)

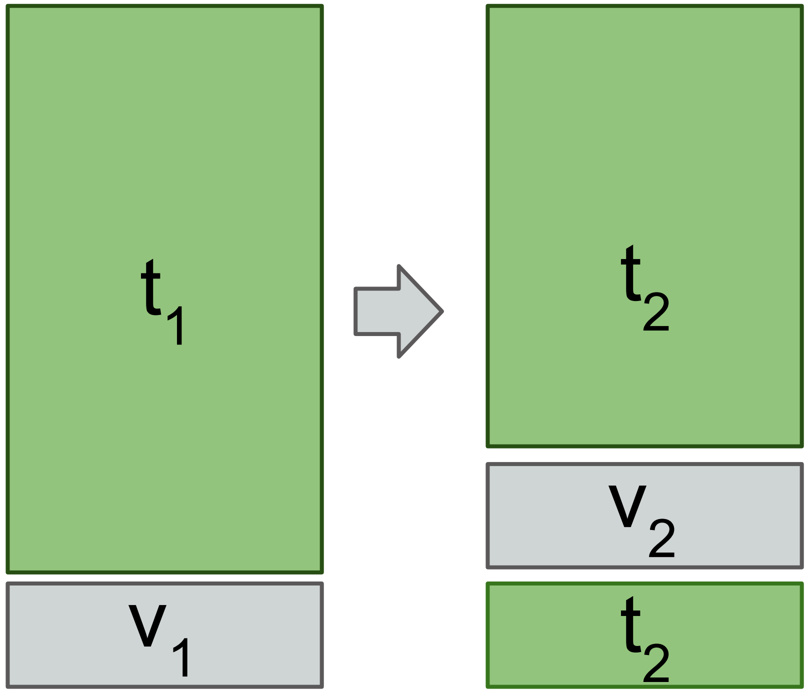
Methods - Coding algorithmically

- Validation
  - Assessing accuracy on test data is misleading, hence the need for validation
  - We use stratified, 10-fold **cross validation**



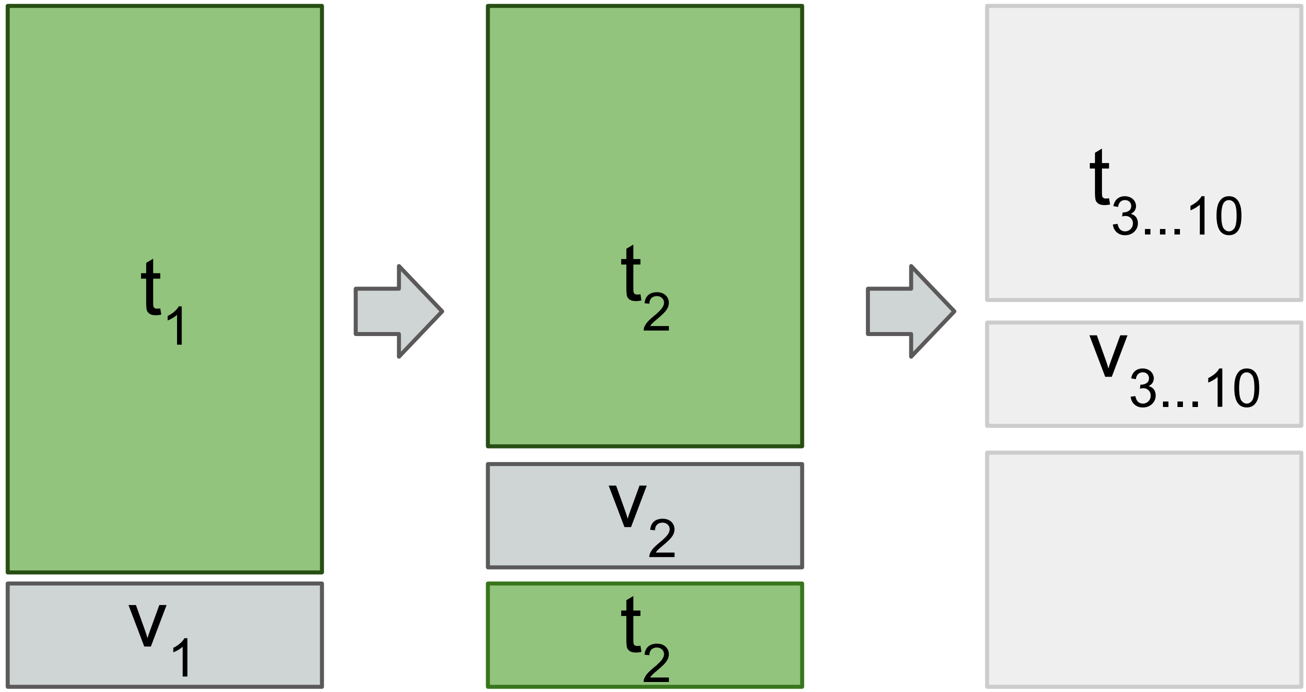
Methods - Coding algorithmically

- Validation
  - Assessing accuracy on test data is misleading, hence the need for validation
  - We use stratified, 10-fold cross validation



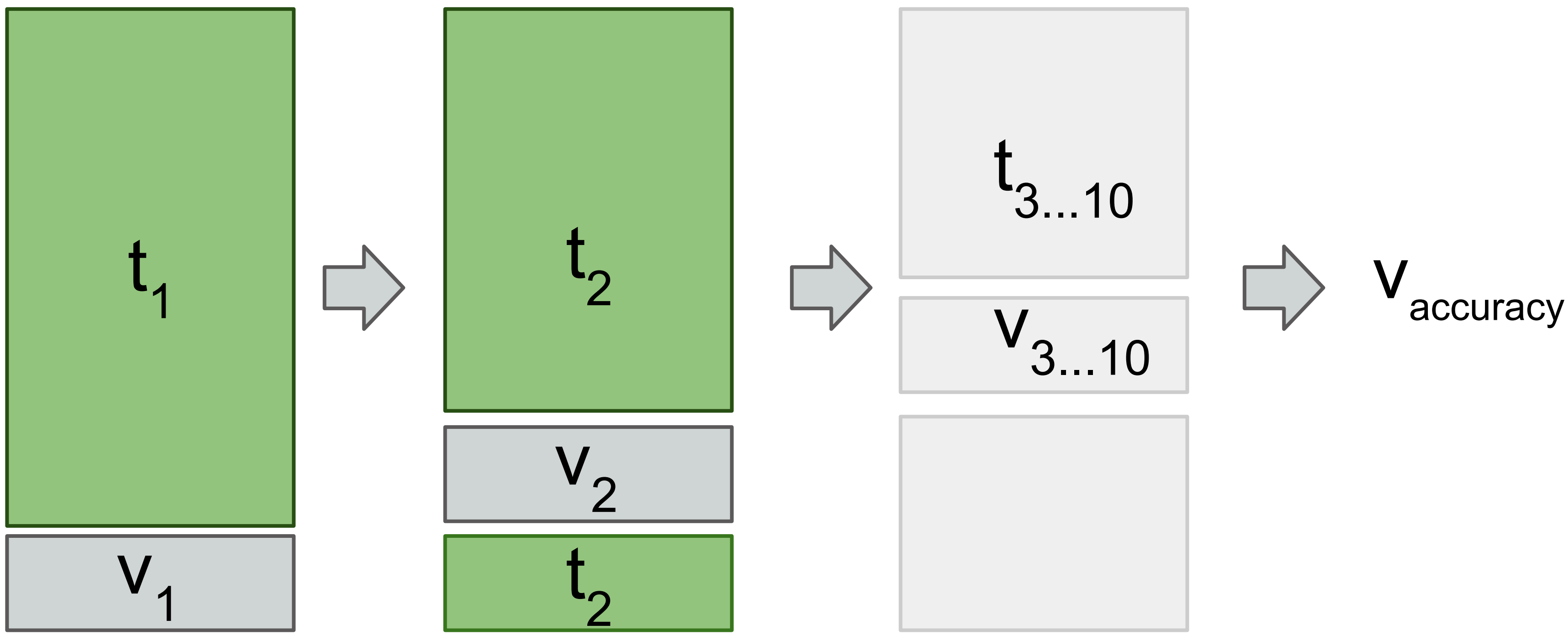
Methods - Coding algorithmically

- Validation
  - Assessing accuracy on test data is misleading, hence the need for validation
  - We use stratified, 10-fold cross validation



Methods - Coding algorithmically

- Validation
  - Assessing accuracy on test data is misleading, hence the need for validation
  - We use stratified, 10-fold cross validation



Methods - Coding algorithmically

- **One-versus-all technique**
  - Transforms a multi-class classification task into n-binary classifications tasks
  - Typically used with Support Vector Machines, but produces some results with naive Bayes classifiers

(Sulzmann, Fürnkranz, & Hüllermeier, 2007)

Methods - Coding algorithmically

- One-versus-all technique

$tweet\_text_1$ $class_{[1:5]}$

$tweet\_text_2$ $class_{[1:5]}$

$tweet\_text_3$ $class_{[1:5]}$

Methods - Coding algorithmically

- One-versus-all technique

tweet_text_1 $\text{class}_{[1:5]}$

tweet_text_2 $\text{class}_{[1:5]}$

tweet_text_3 $\text{class}_{[1:5]}$

tweet_text_1 $\text{class\_}\mathbf{1}?_{[0,1]}$

tweet_text_1 $\text{class\_}\mathbf{2}?_{[0,1]}$

tweet_text_1 $\text{class\_}\mathbf{3}?_{[0,1]}$

tweet_text_1 $\text{class\_}\mathbf{4}?_{[0,1]}$

tweet_text_1 $\text{class\_}\mathbf{5}?_{[0,1]}$

Methods - Coding algorithmically

- One-versus-all technique

tweet_text_1 $\text{class}_{[1:5]}$

tweet_text_2 $\text{class}_{[1:5]}$

tweet_text_3 $\text{class}_{[1:5]}$

tweet_text_2 $\text{class\_}\mathbf{1}?_{[0,1]}$

tweet_text_2 $\text{class\_}\mathbf{2}?_{[0,1]}$

tweet_text_2 $\text{class\_}\mathbf{3}?_{[0,1]}$

tweet_text_2 $\text{class\_}\mathbf{4}?_{[0,1]}$

tweet_text_2 $\text{class\_}\mathbf{5}?_{[0,1]}$

Methods - Coding algorithmically

- One-versus-all technique

tweet_text_1 $\text{class}_{[1:5]}$

tweet_text_2 $\text{class}_{[1:5]}$

tweet_text_3 $\text{class}_{[1:5]}$

tweet_text_3 $\text{class\_}\mathbf{1}?_{[0,1]}$

tweet_text_3 $\text{class\_}\mathbf{2}?_{[0,1]}$

tweet_text_3 $\text{class\_}\mathbf{3}?_{[0,1]}$

tweet_text_3 $\text{class\_}\mathbf{4}?_{[0,1]}$

tweet_text_3 $\text{class\_}\mathbf{5}?_{[0,1]}$