

Detecting Information Need in Social Media

Zhe Zhao

Outline

- Motivation
- Problem Description
- Our Solution
- Experiment results
- Challenges & Discussion
- Current Status & Plan

Information Need?

- What is information need?
 - Direct Explanation:
 - The need to know something
 - WIKI
 - An individual or group's desire to locate and obtain information to satisfy a conscious or unconscious need.

Information Need?

- Examples:
 - What is Support Vector Machine?
 - How to use MATLAB to read files?
 - Where is Jim Gray?
 - If I want to adopt a dog, what do I need to do?
 - Where would we go to lunch?
- What do we do if we have information need?

How to find information

- Search?



AnonID	Query	QueryTime	ItemRank	ClickURL
100218	tennessee department of transportation	2006-03-01 11:08:30	1	http://www.tdot.state.tn.us
100218	tennessee federal court	2006-03-01 11:53:44	1	http://www.constructionweblinks.com
100218	state of tennessee emergency communications board	2006-03-01 12:56:18	1	http://www.tennessee.gov
100218	state of tennessee emergency communications board	2006-03-01 12:56:18	1	http://www.tennessee.gov
100218	state of tennessee emergency communications board	2006-03-01 12:56:18	2	http://www.tennessee.gov
100218	state of tennessee emergency communications board	2006-03-01 12:56:18	1	http://www.tennessee.gov
100218	dixie youth softball	2006-03-02 10:36:48	2	http://www.dixie.org
100218	cdwg	2006-03-03 14:29:07	1	http://www.cdwg.com
100218	cdwg scam cdwge	2006-03-03 14:30:11		
100218	escambia county sheriff's department	2006-03-07 09:26:51	1	http://www.escambiaso.com
100218	escambia county sheriff's department	2006-03-07 09:26:51	2	http://www.escambiaso.com
100218	escambia county sheriff's department	2006-03-07 09:26:51	1	http://www.escambiaso.com
100218	escambia county sheriff's department	2006-03-07 09:26:51	1	http://www.escambiaso.com
100218	pensacola police department	2006-03-07 09:34:28	1	http://www.pensacolapolice.com
100218	memphis pd	2006-03-07 09:42:33	1	http://www.memphispolice.org
100218	nashville metro pd	2006-03-07 09:44:43	1	http://www.police.nashville.org
100218	florida highway patrol	2006-03-07 09:48:35	1	http://www.fhp.state.fl.us
100218	tennessee highway patrol	2006-03-07 09:49:52	1	http://www.state.tn.us
100218	florida bureau of investigations	2006-03-07 09:51:08	2	http://www.flbsi.com
100218	florida bureau of investigations	2006-03-07 09:51:08	1	http://www.fhp.state.fl.us
100218	government finance officers association	2006-03-07 21:16:11		
100218	state of tennessee controllers manual	2006-03-07 21:17:12		
100218	state of tennessee audit controllers manual	2006-03-07 21:17:40	3	http://www.comptroller.state.tn.us
100218	state of tennessee audit controllers manual	2006-03-07 21:17:40	4	http://www.fbr.state.tn.us
100218	state of tennessee audit controllers manual	2006-03-07 21:17:40	9	http://audit.tennessee.edu
100218	internal controls for municipalities under 10 000	2006-03-07 21:38:04	1	http://www.nysscpa.org
100218	internal controls for municipalities under 10 000	2006-03-07 21:38:04	4	http://www.massdor.com
100218	municipality fraud detection techniques	2006-03-07 21:41:40		
100218	municipal fraud audit detection internal controls	2006-03-07 21:43:15		
100218	internal fraud controls for municipalities cities towns local government	2006-03-07 21:45:13	1	http://www.whitehouse.gov
100218	internal fraud controls for municipalities cities towns local government	2006-03-07 21:45:13	4	http://www.nhlgc.org
100218	internal fraud controls for municipalities cities towns local government	2006-03-07 21:45:13	7	http://www.sao.state.ut.us
100218	evaluating internal controls a local government managers guide	2006-03-07 21:51:18	5	http://www.allbusiness.com

How to find information

- Search?



1. Formulate our Information Need in Keywords.
2. Whether you can get information depend on how you search

AnonID	Query	QueryTime	ItemRank	ClickURL
100218	tennessee department of transportation	2006-03-01 11:08:30	1	http://www.tdot.state.tn.us
100218	tennessee federal court	2006-03-01 11:53:44	1	http://www.constructionweblinks.com
100218	state of tennessee emergency communications board	2006-03-01 12:56:18	1	http://www.tennessee.gov
100218	state of tennessee emergency communications board	2006-03-01 12:56:18	1	http://www.tennessee.gov
100218	state of tennessee emergency communications board	2006-03-01 12:56:18	2	http://www.tennessee.gov
100218	state of tennessee emergency communications board	2006-03-01 12:56:18	1	http://www.tennessee.gov
100218	dixie youth softball	2006-03-02 10:36:48	2	http://www.dixie.org
100218	cdwg	2006-03-03 14:29:07	1	http://www.cdwg.com
100218	cdwg scam cdwge	2006-03-03 14:30:11		
100218	escambia county sheriff's department	2006-03-07 09:26:51	1	http://www.escambiaso.com
100218	escambia county sheriff's department	2006-03-07 09:26:51	2	http://www.escambiaso.com
100218	escambia county sheriff's department	2006-03-07 09:26:51	1	http://www.escambiaso.com
100218	escambia county sheriff's department	2006-03-07 09:26:51	1	http://www.escambiaso.com
100218	pensacola police department	2006-03-07 09:34:28	1	http://www.pensacolapolice.com
100218	memphis pd	2006-03-07 09:42:33	1	http://www.memphispolice.org
100218	nashville metro pd	2006-03-07 09:44:43	1	http://www.police.nashville.org
100218	florida highway patrol	2006-03-07 09:48:35	1	http://www.fhp.state.fl.us
100218	tennessee highway patrol	2006-03-07 09:49:52	1	http://www.state.tn.us
100218	florida bureau of investigations	2006-03-07 09:51:08	2	http://www.flbsi.com
100218	florida bureau of investigations	2006-03-07 09:51:08	1	http://www.fhp.state.fl.us
100218	government finance officers association	2006-03-07 21:16:11		
100218	state of tennessee controllers manual	2006-03-07 21:17:12		
100218	state of tennessee audit controllers manual	2006-03-07 21:17:40	3	http://www.comptroller.state.tn.us
100218	state of tennessee audit controllers manual	2006-03-07 21:17:40	4	http://www.fbr.state.tn.us
100218	state of tennessee audit controllers manual	2006-03-07 21:17:40	9	http://audit.tennessee.edu
100218	internal controls for municipalities under 10 000	2006-03-07 21:38:04	1	http://www.nysscpa.org
100218	internal controls for municipalities under 10 000	2006-03-07 21:38:04	4	http://www.massdor.com
100218	municipality fraud detection techniques	2006-03-07 21:41:40		
100218	municipal fraud audit detection internal controls	2006-03-07 21:43:15		
100218	internal fraud controls for municipalities cities towns local government	2006-03-07 21:45:13	1	http://www.whitehouse.gov
100218	internal fraud controls for municipalities cities towns local government	2006-03-07 21:45:13	4	http://www.nhlgc.org
100218	internal fraud controls for municipalities cities towns local government	2006-03-07 21:45:13	7	http://www.sao.state.ut.us
100218	evaluating internal controls a local government managers guide	2006-03-07 21:51:18	5	http://www.allbusiness.com

How to find information

- Search? **Ask?**

YAHOO! ANSWERS



Who sang the doing the bump song?

☆ In Other - Music - Asked by Tremaine - 0 answers - 2 seconds ago



Question about whether a transformation is linear?

☆ In Mathematics - Asked by Someone - 0 answers - 5 seconds ago



The change in eye color due to hair shade?

☆ In Other - Beauty & Style - Asked by Alyssa - 0 answers - 8 seconds ago



Which telescope is better?

☆ In Astronomy & Space - Asked by hildy - 0 answers - 10 seconds ago



Problem with xbox live marketplace?

☆ In Xbox - Asked by liam - 0 answers - 12 seconds ago



Online site for submitting articles?

☆ In Other - Careers & Employment - Asked by Shahnaz - 0 answers - 14 seconds ago



What is the best way to teach your children about the Holocaust?

☆ In History - Asked by Comrade Good Looking - 0 answers - 16 seconds ago

How to find information

- Search? **Ask?**

1. Have to expect some random guy to answer.
2. Not always possible to get answers

YAHOO! ANSWERS



Who sang the doing the bump song?

☆ In Other - Music - Asked by Tremaine - 0 answers - 2 seconds ago



Question about whether a transformation is linear?

☆ In Mathematics - Asked by Someone - 0 answers - 5 seconds ago



The change in eye color due to hair shade?

☆ In Other - Beauty & Style - Asked by Alyssa - 0 answers - 8 seconds ago



Which telescope is better?

☆ In Astronomy & Space - Asked by hildy - 0 answers - 10 seconds ago



Problem with xbox live marketplace?

☆ In Xbox - Asked by liam - 0 answers - 12 seconds ago



Online site for submitting articles?

☆ In Other - Careers & Employment - Asked by Shahnaz - 0 answers - 14 seconds ago



What is the best way to teach your children about the Holocaust?

☆ In History - Asked by Comrade Good Looking - 0 answers - 16 seconds ago

How to find information

- Search? Ask? Or **Tweet?**



K4m74n Kamran Hussain

"@serenaakellyy: So who **knows** some good horror movies? Anyone at all?" Grudge 1&2 #soscary

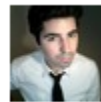
6 minutes ago



funnysg Funny Singapore

Anyone **knows** of any REALISTIC ways to make money for a teenager? Please mention me!

30 Nov



felixcartal FELIX CARTAL

What is everyone's opinions on mittens? #ImportantQuestions

2 hours ago



eduardo_moran Eduardo Morán

Where have you been, cause I never see you out, are you hiding from me, yeah?

4 hours ago



sajha sajha

bholebom: Does anybody has **carfax** I would appreciate if you can run this VIN number 1HCG325XXA023312 sajha.co/?93602 #nepal

36 minutes ago



POPtartsNPornos Ebony Ayes

I really **need** a new **matress** but I'm to cheap to save money to actually buy one....

20 hours ago

How to find information

• Search? Ask? Or **Tweet?**



1. We can ask questions for varies of information
2. Personalized Answer can be provided by friends
3. Friends would be more likely to answer if they know.
4. If they can't answer, they retweet, and broadcast the need.



K4m74n Kamran Hussain

"@serenaakellyy: So who **knows** some good horror movies? **Anyone** at all?" Grudge 1&2 #soscary

6 minutes ago



funnysg Funny Singapore

Anyone knows of any REALISTIC ways to make money for a teenager? Please mention me!

30 Nov



felixcartal FELIX CARTAL

What is everyone's opinions on mittens? #ImportantQuestions

2 hours ago



eduardo_moran Eduardo Morán

Where have you been, cause I never see you out, are you hiding from me, yeah?

4 hours ago



sajha sajha

bholebom: Does anybody has **carfax** I would appreciate if you can run this VIN number 1HCG325XXA023312 sajha.co/?93602 #nepal

36 minutes ago



POPtartsNPornos Ebony Ayes

I really **need** a new **matress** but I'm to cheap to save money to actually buy one....

20 hours ago

Our Motivation

- Information Need in Social Media
 - Much more complicated and richer than search engine and Q&A
 - Study Information Need in Social Media is important in study social media and new forms of web services.
- Detecting Information Need in Social Media
 - Itself is meaningful
 - Detect Trend of Information Need
 - To fulfill information need is even more meaningful

Problem Description

- Detect user's information need in social media
 - Given User's generated content, and his/her profile(behaviors), detect whether the content he/she generated represent information need
- Formulation
 - Input: social media content feature and extra information.
 - Output: 1 if represent information need, 0 if not.

Our Solution

- How can we detect Information Need on twitter?



Our Solution

- How can we detect Information Need on twitter?
 - Train a binary classifier.



Our Solution

- How can we detect Information Need on twitter?
 - Train a binary classifier.
 - Label some data
 - Train an SVM



Our Solution

- How can we detect Information Need on twitter?
 - Train a binary classifier.
 - Label some data
 - Train an SVM
 - Seems alright...



Our Solution

- How can we detect Information Need on twitter?
 - Train a binary classifier.
 - Label some data
 - Train an SVM
 - Seems alright...
- How's this even hard?



Our Solution

- How can we detect Information Need on twitter?
 - Train a binary classifier.
 - Label some data
 - Train an SVM
 - Seems alright...
- How's this even hard?



Our Solution

- How can we detect Information Need on twitter?
 - Train a binary classifier.
 - Label some data
 - Train an SVM
 - Seems alright...
- How's this even hard?
 - Limited text feature in tweets
 - Tweets from varies of topic, filled with noise
 - Rich noisy surrounding features
 - Tweets with IN in all tweets may be less than 5-10%



Our Solution

- Two challenges
 - Lack of Content Feature
 - Lack of Labeled data

Our Solution

- Two challenges
 - Lack of Content Feature
- social media is more than this.

Our Solution

- Two challenges
 - Lack of Content Feature
- social media is more than this.


 **OccupyChicago** Occupy Chicago
We **need a car**, van, or SUV asap at Jackson and Lasalle to transport supplies 6 blocks. Call T, 312-502-6086
5 Oct ☆ Favorite ↻ Retweet ↩ Reply

Retweeted by [ZombiesAreBad](#) and 23 others



replies ↓

 **SetBrushFires** John Q. Public
[@OccupyChicago](#) just make sure it is a Ford SUV, they didn't take the bailout money, so its ok if theory SUV's pollute!!!
5 Oct

 **BruceEParry** Bruce Parry
[@OccupyChicago](#) I have a Camry. Bruce at 773-320-1859.
5 Oct

 **SetBrushFires** John Q. Public
[@OccupyChicago](#) SUV's hurt mother earth!!!!
5 Oct

Our Solution

- Two challenges
 - Lack of Content Feature

- social media is more than this.



OccupyChicago Occupy Chicago
We **need a car**, van, or SUV asap at Jackson and Lasalle to transport supplies 6 blocks. Call T, 312-502-6086
5 Oct ☆ Favorite ↻ Retweet ↩ Reply

Retweeted by [ZombiesAreBad](#) and 23 others



replies ↓

SetBrushFires John Q. Public
@OccupyChicago just make sure it is a Ford SUV, they didn't take the bailout money, so its ok if theory SUV's pollute!!!
5 Oct

BruceEParry Bruce Parry
@OccupyChicago I have a Camry. Bruce at 773-320-1859.
5 Oct

SetBrushFires John Q. Public
@OccupyChicago SUV's hurt mother earth!!!!
5 Oct

Our Solution

- Feature Interaction Graph(FIG)[2]
 - Markov Random Field in Similarity Measure[1,2]

Nodes:

Feature: Each feature of the objects

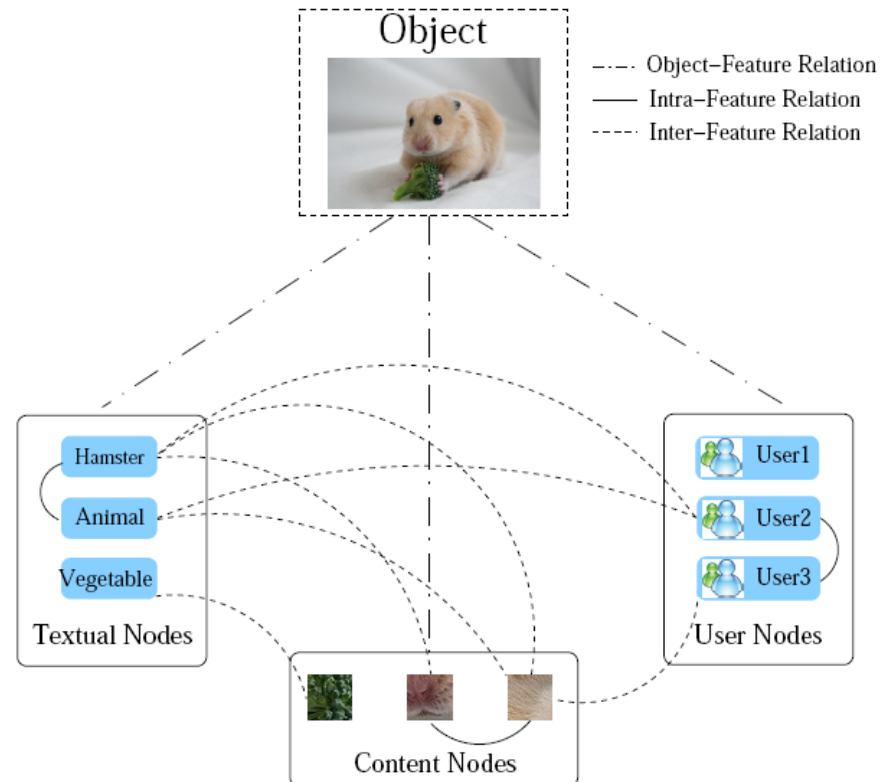
Object: Virtual Root Node

Edges:

Object to Feature

Feature to Feature

Intra type: Similarity between each feature
Inter type: Correlation between different type of features.



Our Solution

- Feature Interaction Graph(FIG)[2]
 - Markov Random Field in Similarity Measure[1,2]

Nodes:

Feature: Each feature of the objects

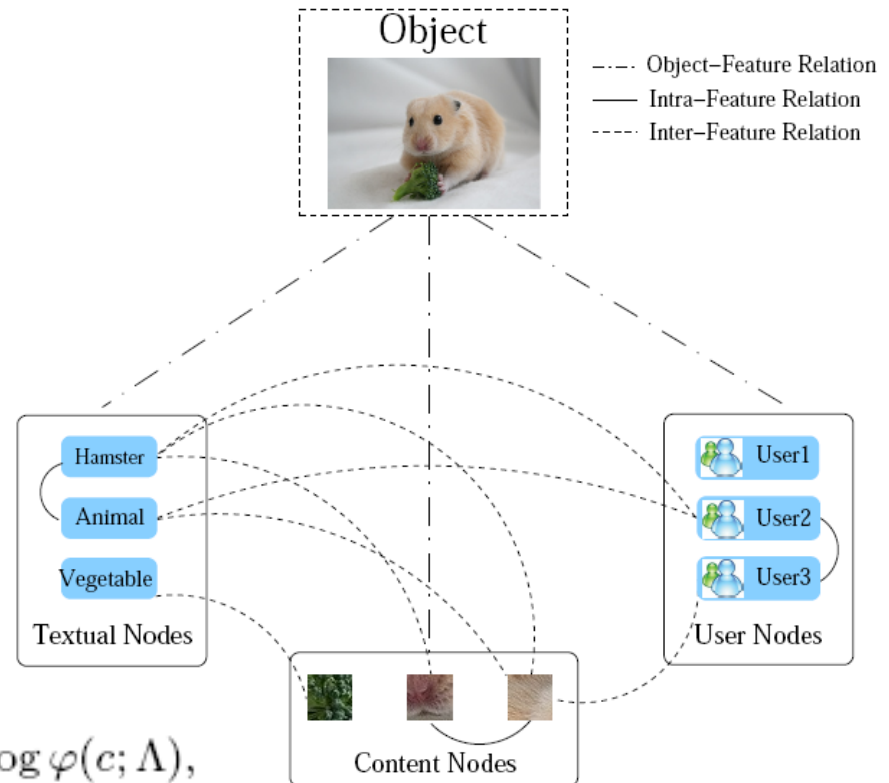
Object: Virtual Root Node

Edges:

Object to Feature

Feature to Feature

Intra type: Similarity between each feature
 Inter type: Correlation between different type of features.



$$P(G') = \frac{1}{Z_{\Lambda}} \prod_{c \in C(G')} \varphi(c; \Lambda) \propto \sum_{c \in C(G')} \log \varphi(c; \Lambda),$$

Our Solution

- Feature Interaction Graph(FIG)[2]
 - Markov Random Field in Similarity Measure[1,2]

Nodes:

Feature: Each feature of the objects

Object: Virtual Root Node

Edges:

Object to Feature

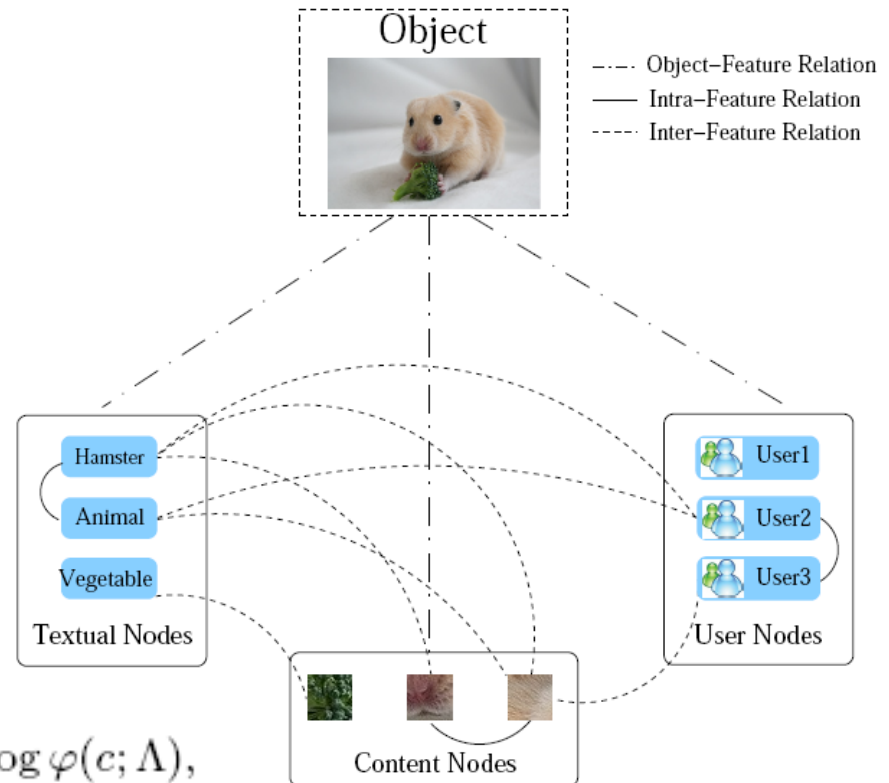
Feature to Feature

Intra type: Similarity between each feature
 Inter type: Correlation between different type of features.

Kernel Function



$$P(G') = \frac{1}{Z_{\Lambda}} \prod_{c \in C(G')} \varphi(c; \Lambda) \propto \sum_{c \in C(G')} \log \varphi(c; \Lambda),$$



Our Solution

- Two challenges
 - Lack of Labeled data

Our Solution

- Two challenges
 - Lack of Labeled data
- Train multiple classifiers in different subset
 - Subset of questions
 - Train a SVM
 - Subset of statements

Our Solution

- Current Progress:
 - Extract Feature:
 - N-gram
 - Labeled data
 - Tweets with question marks
 - Detected IN
 - On 1M garden hose data from August 2011.

Experiment Results

- Data Sets:
 - Garden Hose data: 10% random sample of twitter stream. ~20M tweets every day.
 - August 1-31, 2011
 - 549.7 M tweets, 61.4M with `?', 104.1M RT, 17M both
 - Crawled data: 100 users' ~25k tweets (in English)
 - 10 selected users and 90 their friends
 - 25840 tweets, 3064 with `?', 705 with selected patterns

What we have done

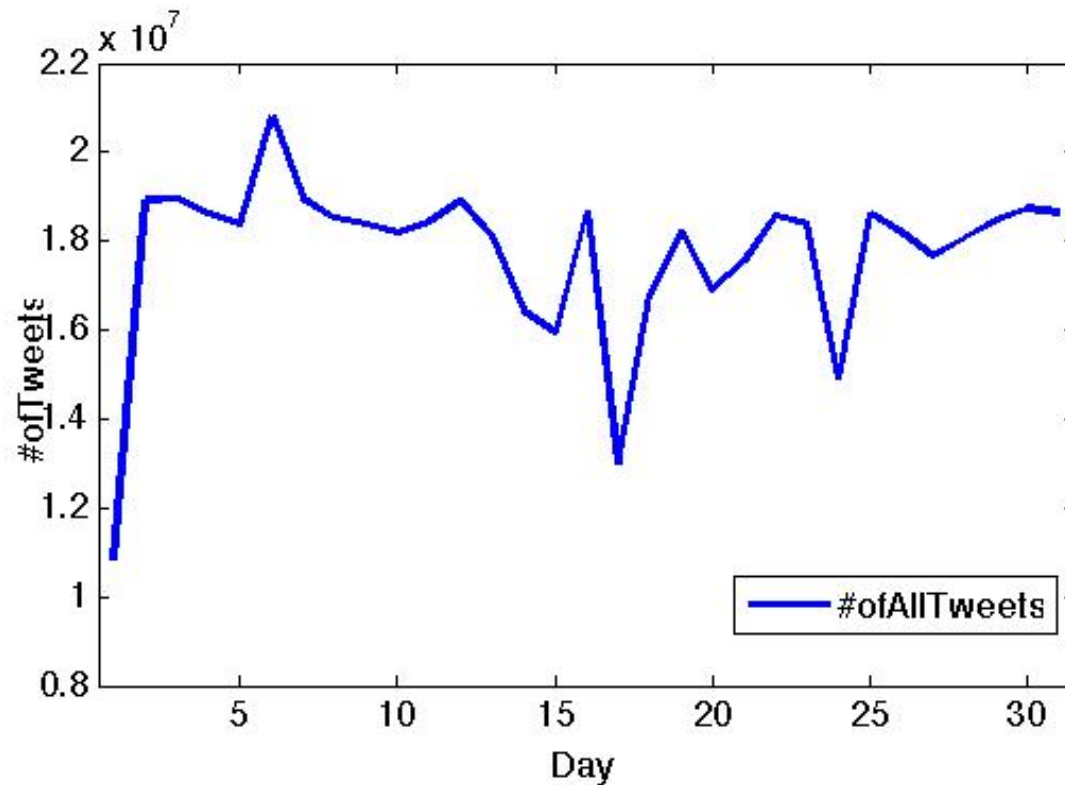
- Study user's asking question and retweeting behavior.
- Study the existence of Information Need in questions.
- Monitor over time
- Train classifier to detect information need
- Tune our model and compare baseline.

Study User Behaviors over Time

- How many tweets are with '?' and RTs in a month?
 - 549.7M tweets
 - 11.2% with '?'
 - 18.9% are retweets
 - 3.1% are both

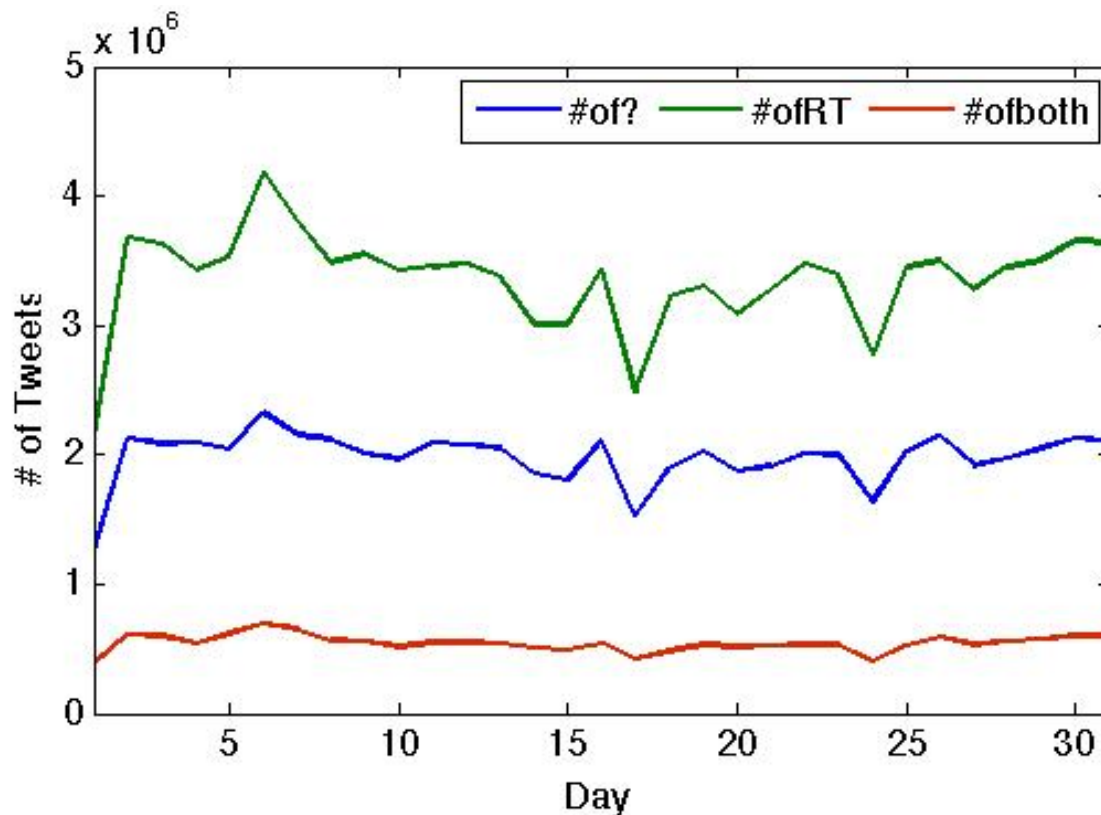
Study User Behaviors over Time

- How many tweets are with '?' and RTs in a month?



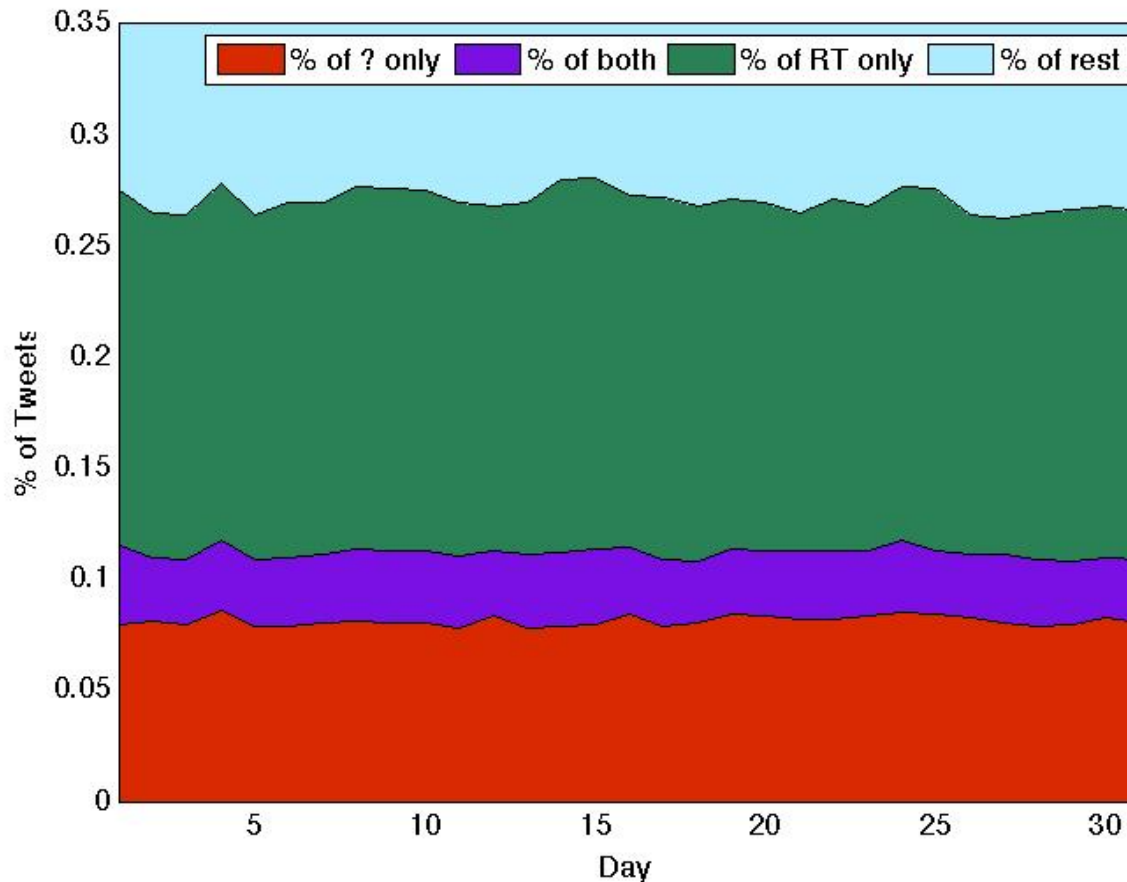
Study User Behaviors over Time

- How many tweets are with '?' and RTs in a month?



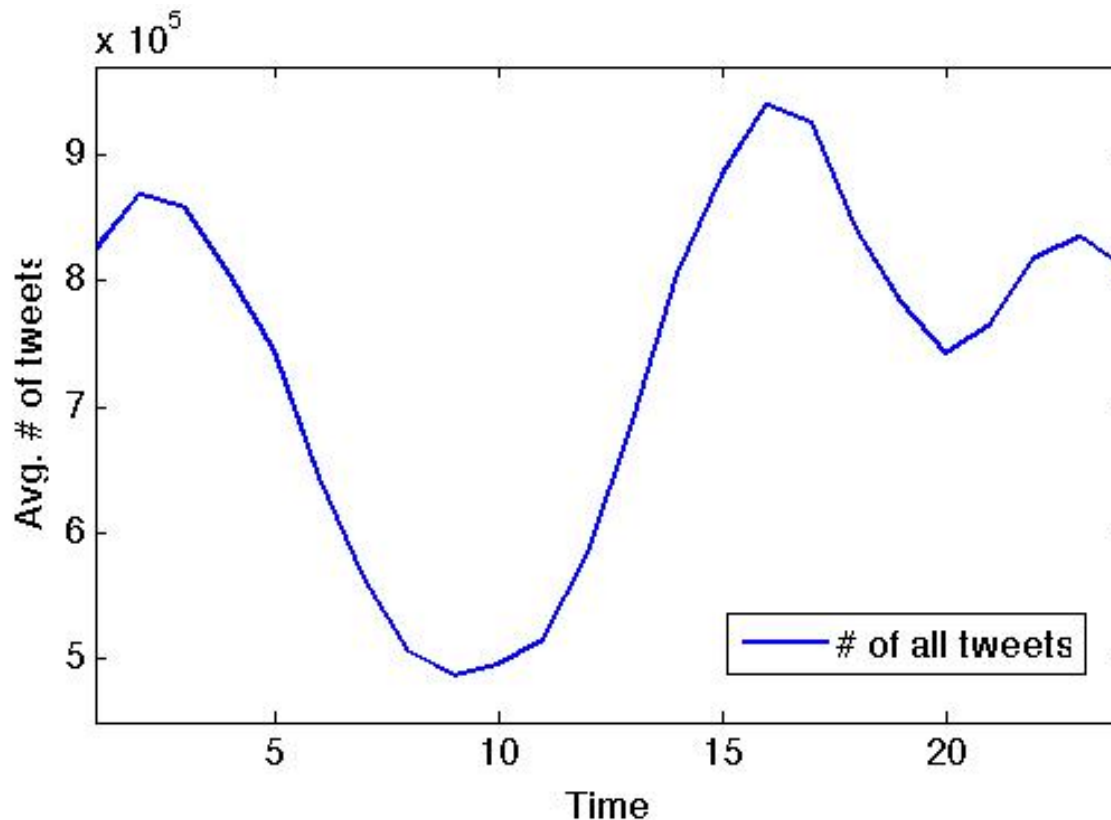
Study User Behaviors over Time

- How many tweets are with '?' and RTs in a month?



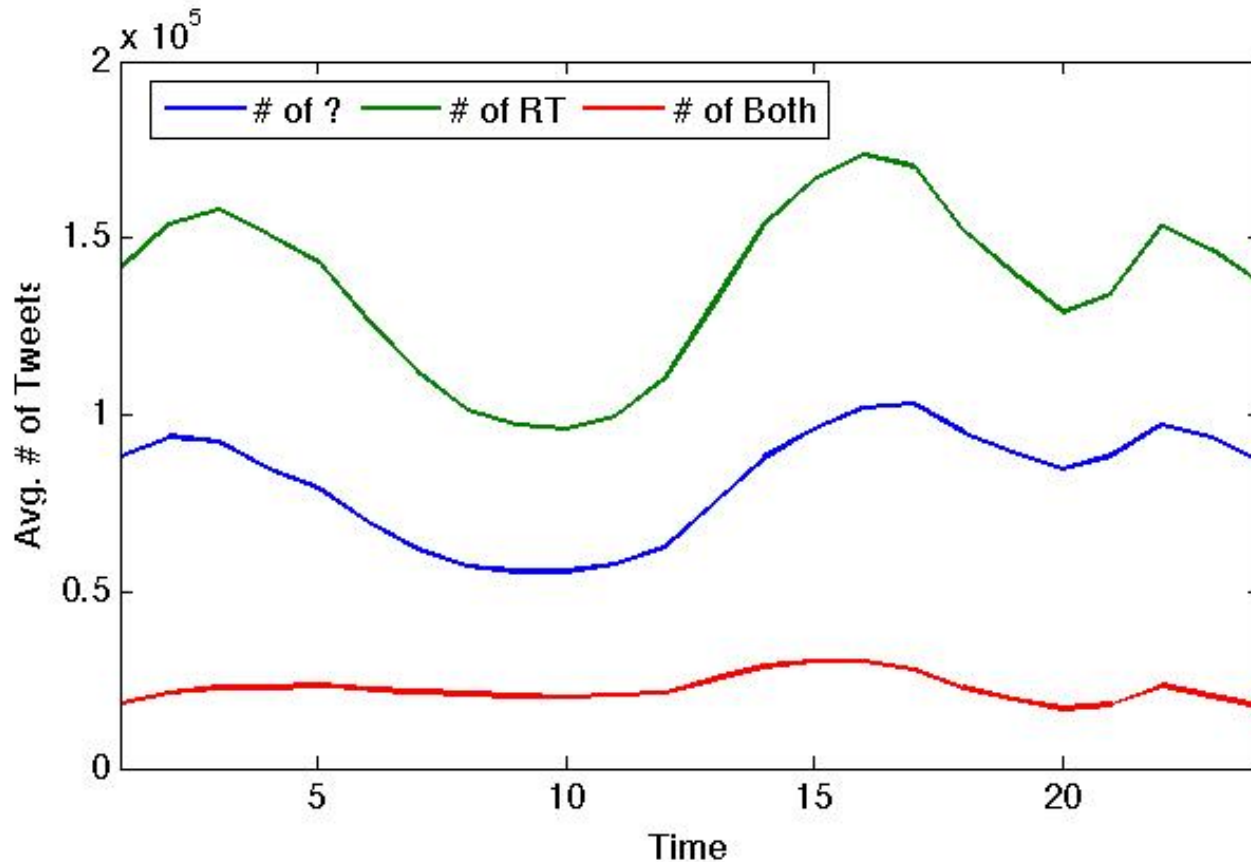
Study User Behaviors over Time

- How many tweets are with '?' and RTs in everyday?



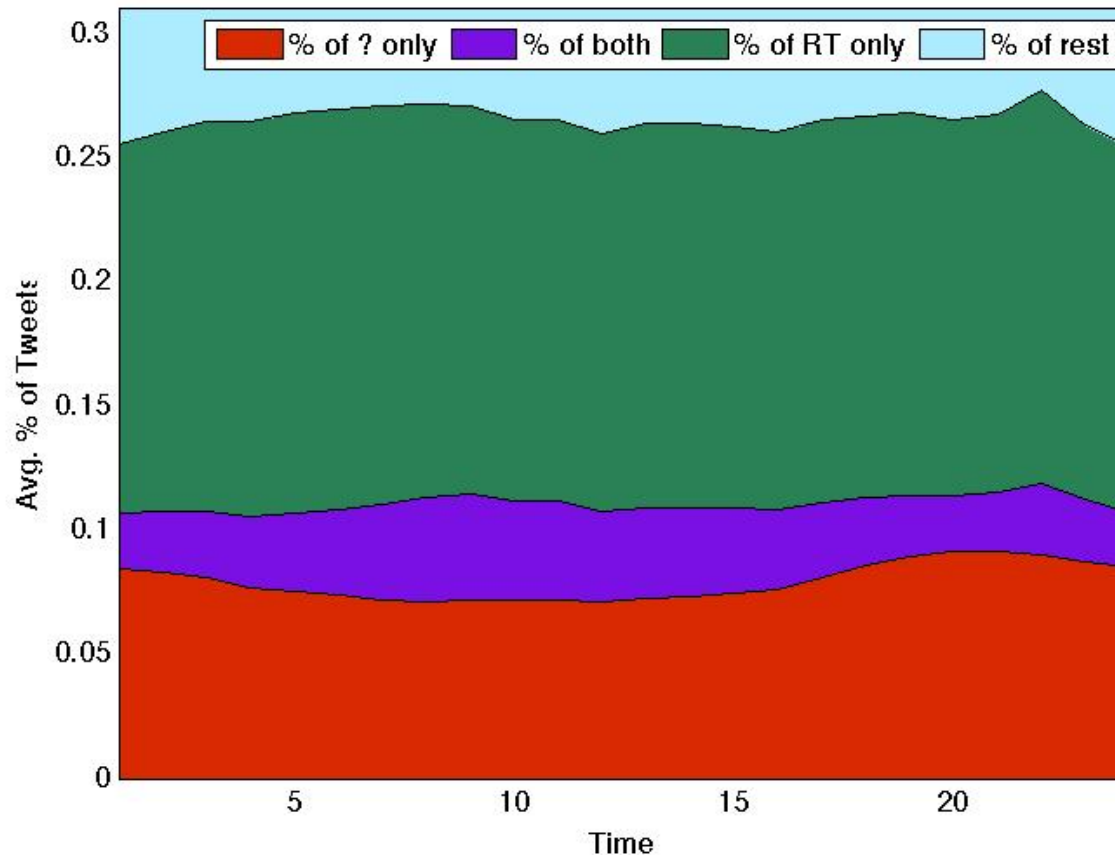
Study User Behaviors over Time

- How many tweets are with '?' and RTs in everyday?



Study User Behaviors over Time

- How many tweets are with '?' and RTs in everyday?



Study the existence of Information Need

- In crawled dataset
 - Select patterns from Q&A websites
 - What should, someone knows, etc..
 - 705 tweets with certain patterns.
 - Label whether tweet represent Information Need
 - 246 tweets represent information needs.
 - 34.9% in tweets with certain patterns
 - Train automatic detector to find Information Need in larger data set.

Study the existence of Information Need

- Classifier Result:
 - Feature Used:
 - 25k n-gram
 - Data Used
 - 360 positive vs. 360 negative
 - 10-fold Cross Validation
 - Precision: 90.3%
- Analysis of the Result:
 - Small Training Set
 - Data only with ? and patterns

Examples in Labeled data

Why does Wilcots think bad things are good? What game is he watching?

How soon will reasonable people really start to think this Northwestern team is ready for the tournament? I'm waiting for conference play.

Where are the replays? Is the FOX truck filled with chimps?

I need help twats.... What would a 4yr old boy want for his birthday from his favorite aunt??

How is CVS out of Vicks??! It's like they want me to feel like shit.

Where is @Kveronica87??!!!!

Does anyone know how common in the past for peers to register their principle residence out of town in order to collect overnight expenses?

Would any of you like to be my henchmen? Job description follows. Strictly hench interns only.

What are we meant to do now? Keep calm and carry on, or run about flapping our hands amusingly and squeaking?

What woman now in England would not be a 'feminist'? We can vote, drive, wear what we like, work and be free to choose. Rest of world?

My hyacinth bulbs are showing above the compost in their bowls. Should I bring them in now? @CatharineHoward

Does anyone know how to restore sounds on my PC which have unaccountably disappeared? Nothing in system tray.

Why won't my computer play sounds any more? I need to watch videos! Help!

What is yr default fashion item that you always find yourself coming home with? Instead of bold new departure you planned when you went out?

It's 2011 - when is the NAACP ever going to change its name? Are they still trying to use up all the old stationery first?

Does anyone know how @nathanallstar is doing ?

Finally moved the basement TV to my room! Success!!! What should I watch while falling asleep?

I just locked our van with the keys inside of it. Haha uh oh what do we do??

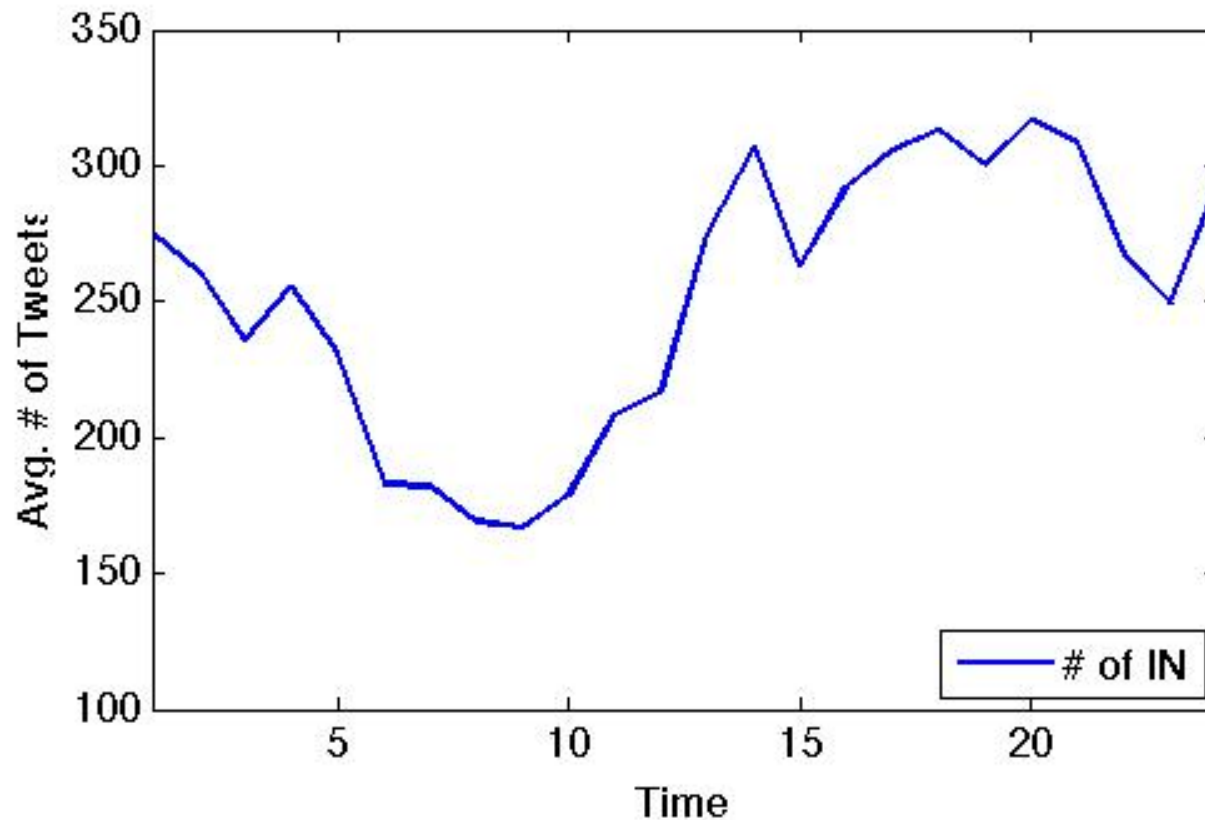
Dennys... Should I get french toast or waffles?

Detecting Information Need In Garden hose dataset

- 0.1% of all the data.

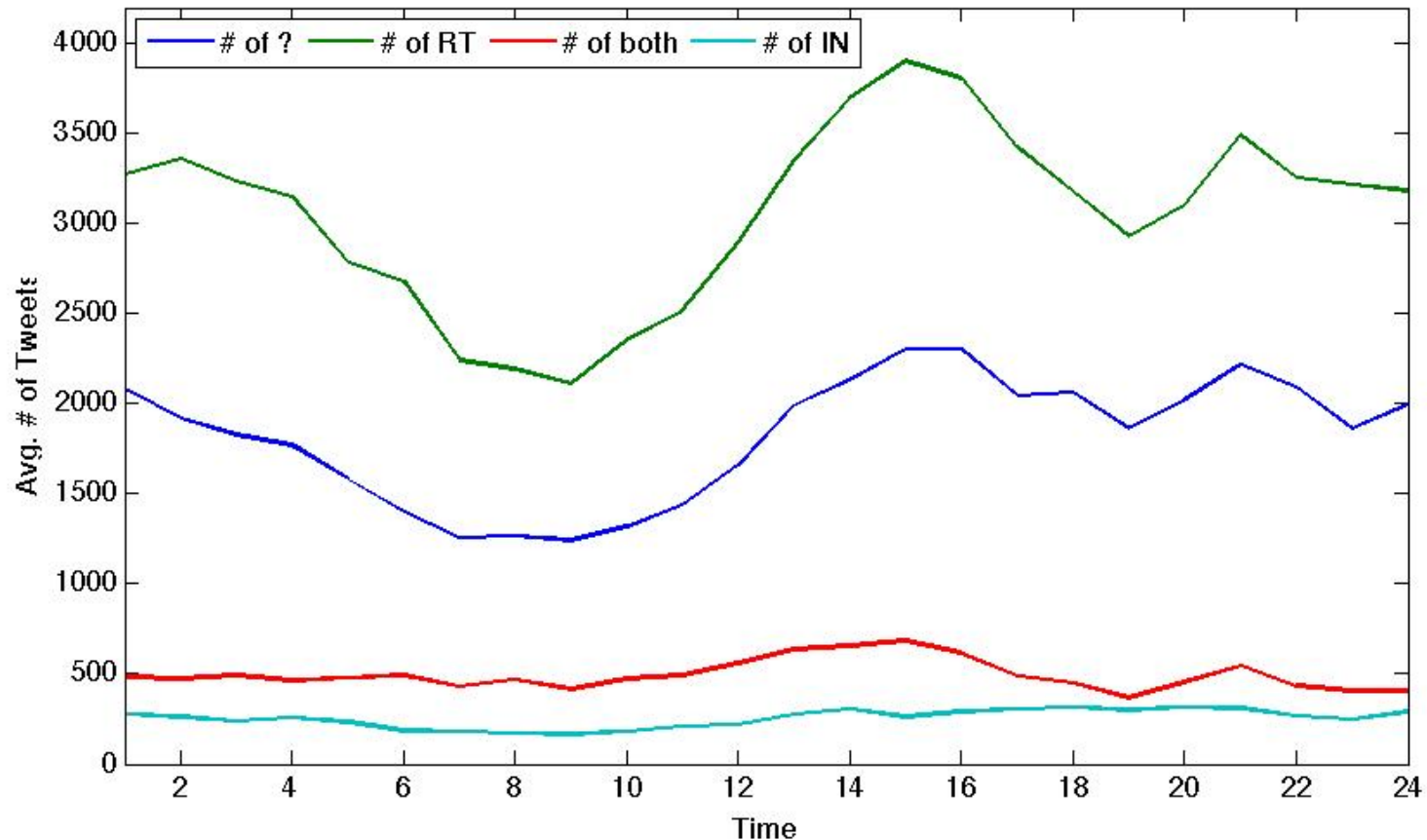
Detecting Information Need In Garden hose dataset

- 0.1% of all the data.



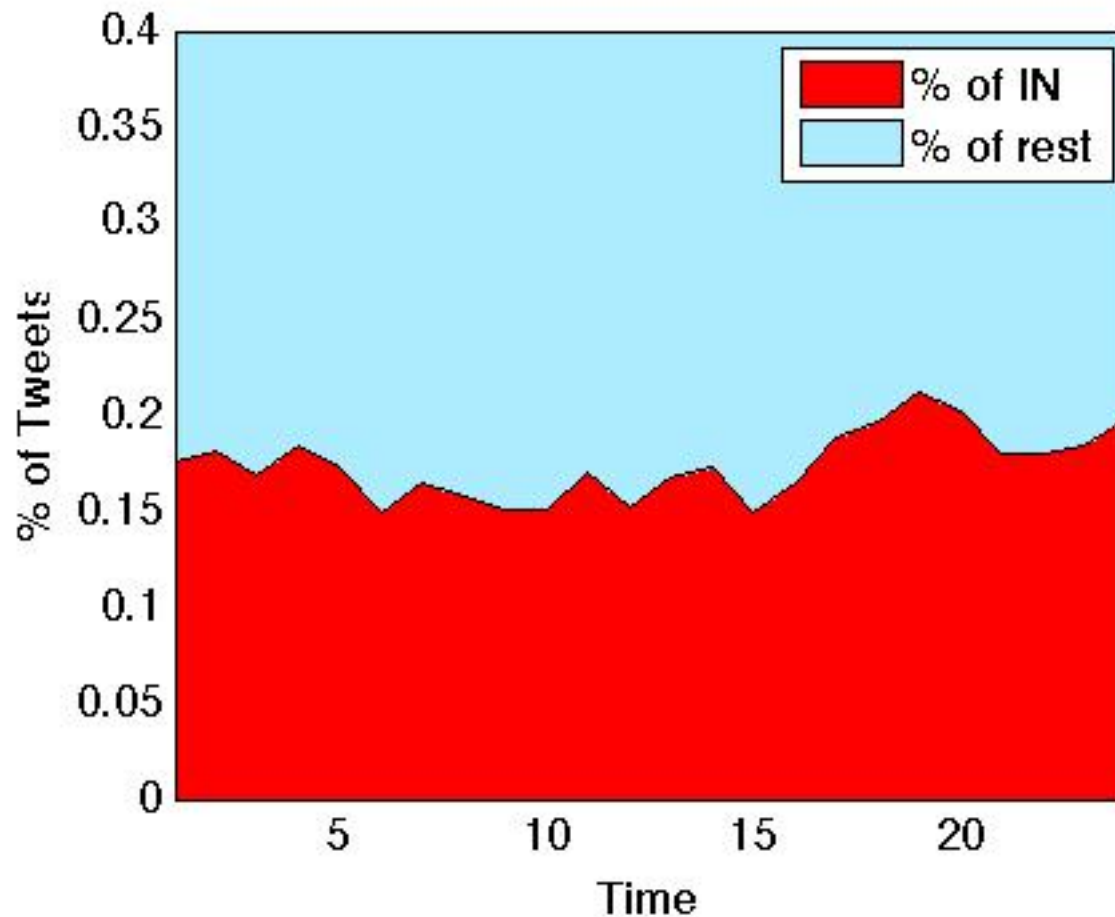
Detecting Information Need In Garden hose dataset

- 0.1% of all the data.

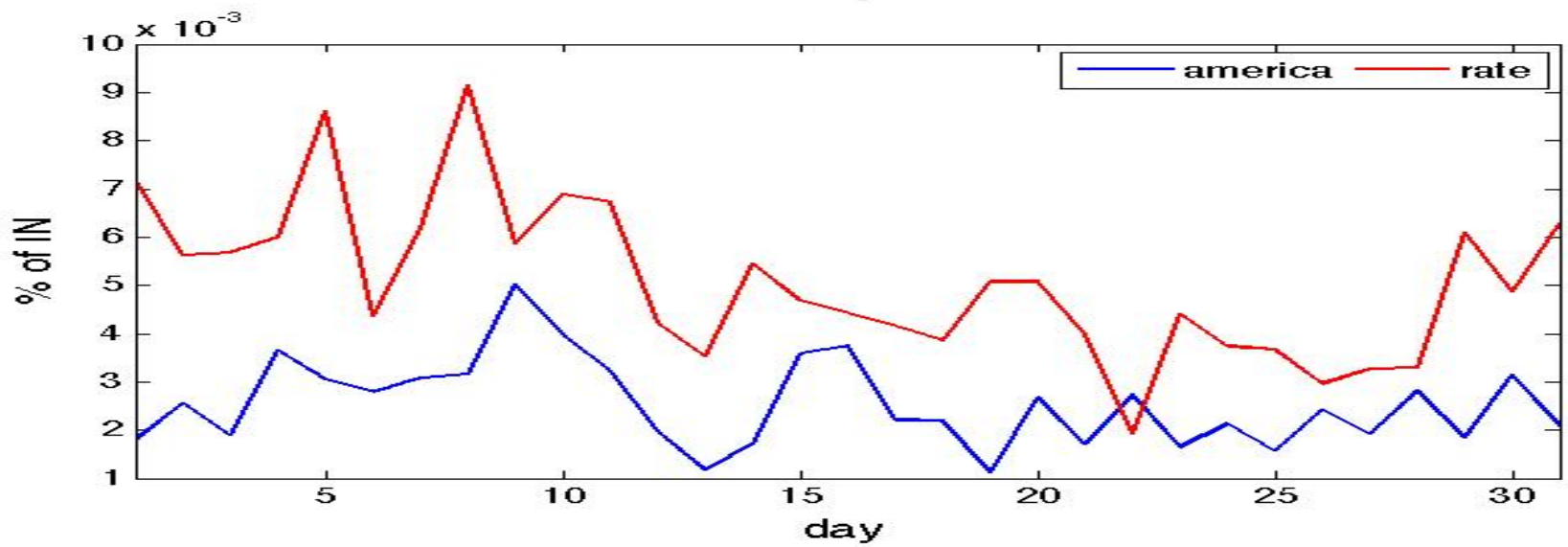
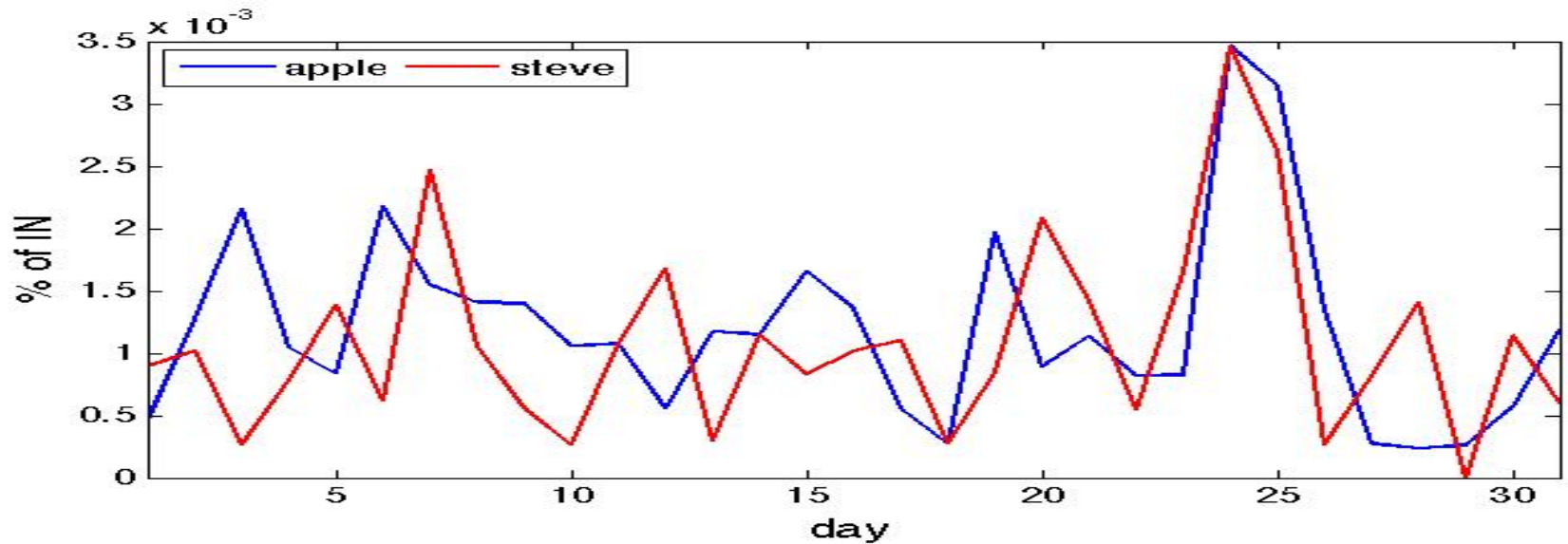


Detecting Information Need In Garden hose dataset

- 0.1% of all the data.



Showing Some Trend



So far...

- Study the Information Need and User behavior in social media
- Extract feature from tweets
- Automatic Detecting Information Need in social media
- Ongoing:
 - Improve classifier for more general tweets
 - Analysis Information Need over time

Thank
you

[1] A Markov random field model for term dependencies. Donald Metzler , W. Bruce Croft,

[2] Multiple feature fusion for social media applications. Bin Cui, Anthony K.H. Tung, Ce Zhang, Zhe Zhao

Challenges

- Make use of other sources of feature
 - Rich feature in social media
 - Social relations and surrounding content
 - How to obtain such data?
- How to detect IN in all tweets?

Challenges

- Large Scalability
 - ~10M tweets per day, ~ 115 tweets per second.
- Work load on each tweets
 - Simplest Case
 - Extract Feature
 - N-gram, maybe in future: Relation feature
 - 80~100 tweets per second single core
 - Classify
 - 40~50 tweets per second single core

Current Status

- Better Classifier
 - NLP to process feature
 - Include extra source of training set
 - Yahoo Answer
 - Include other feature
 - Feature selection (over imbalanced data)
 - More data need to be labeled

Current Status and Plan

- Trend Detection
 - Cluster token into topics
 - Analysis and detect burst and abnormal pattern
- Recommendation
 - Recommend to friend
 - Recommend content
 - Recommend web page

To sum up

- My work is to study user behavior based on Information Need

Thank
you

To sum up

- My work is to study user behavior based on Information Need
- Detect,

Thank
you

To sum up

- My work is to study user behavior based on Information Need
- Detect, Analysis

Thank
you

To sum up

- My work is to study user behavior based on Information Need
- Detect, Analysis and **Make use of** Information Need

Thank
you

To sum up

- My work is to study user behavior based on Information Need
- Detect, Analysis and **Make use of** Information Need
- Long Way to Go:

Thank
you

To sum up

- My work is to study user behavior based on Information Need
- Detect, Analysis and **Make use of** Information Need
- Long Way to Go:
 - Too simple sometimes naïve classifier
 - Lots of feature to explore
 - Look forward to exciting results of analysis

Thank
you