Churn Prediction

Advanced Social Computing

Department of Computer Science
University of Massachusetts, Lowell
Spring 2020

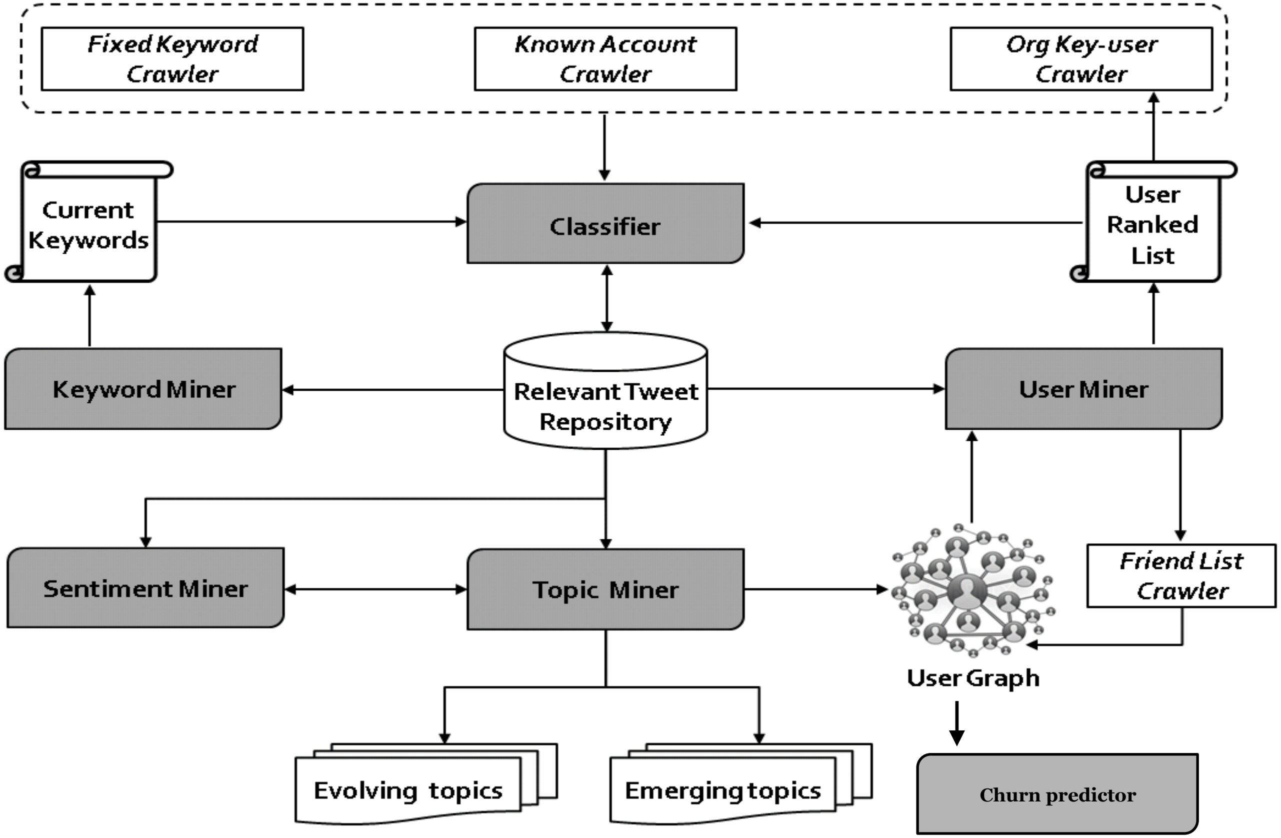
Hadi Amiri
hadi@cs.uml.edu

Lecture Topics

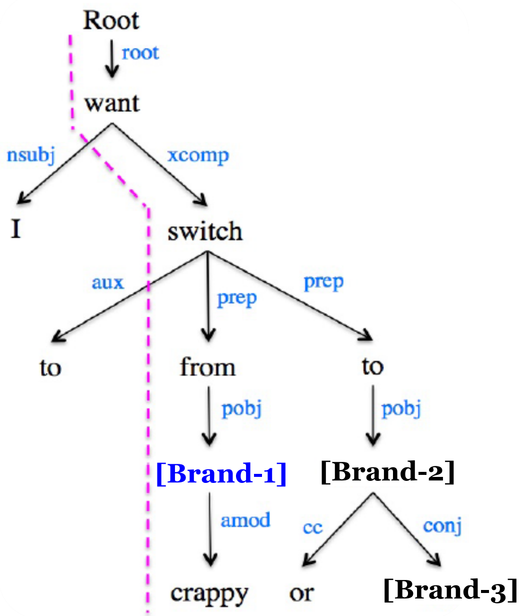
- Churn Prediction

Brandtology

The science of studying brands and their customers in social media.



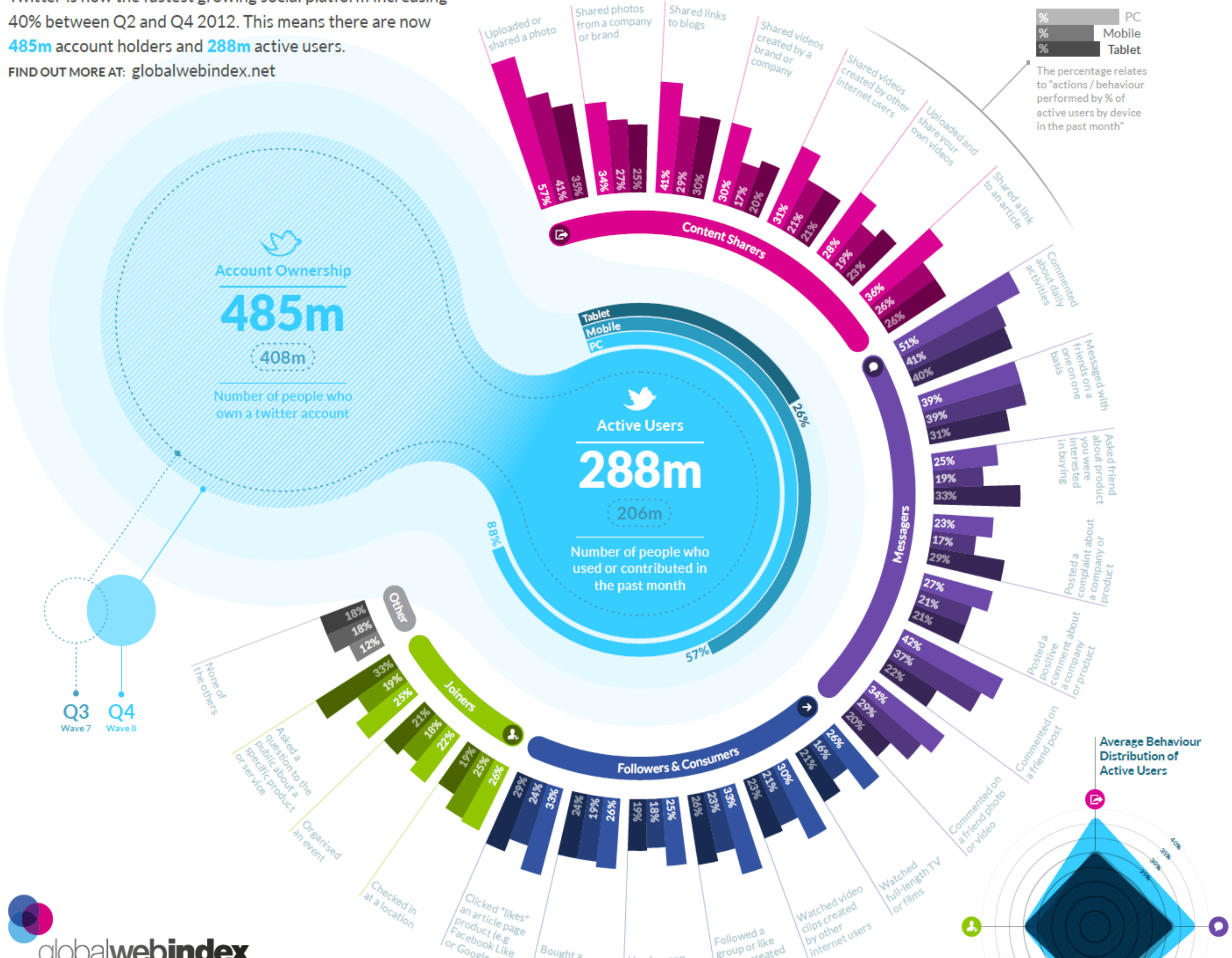
Churn Prediction in Social Media: How can we identify churny content in social media?



TWITTER The Fastest Growing Social Platform

Twitter is now the fastest growing social platform increasing 40% between Q2 and Q4 2012. This means there are now **485m** account holders and **288m** active users.

FIND OUT MORE AT: globalwebindex.net



globalwebindex

The most detailed study on the consumer adoption of the internet ever compiled:

PC /// Mobile /// Tablets /// TV sets /// Gaming
120K+ surveys a year /// 4 waves a year /// 31 markets

Find out more /// www.globalwebindex.net
Contact us /// contact@globalwebindex.net

Twitter Active Users

33% asked friend's **opinion** about a product

26% **bought** a product or service

29% **complained** about a brand or product

27% **recommended** a brand or product

33% **followed** a group created by a brand

25% **used an app** created by a brand

30% **shared photos** about a brand

30% **shared video** created by a brand

Important to extract *insights* form valuable user generated content about brands!

Brands in SM

Sponsored by IBM MARIST

Produced by UNISPHERE RESEARCH Delivering Certainty Thomas J. Wilson, President

Figure 25: Brand and Reputation Monitoring of SMNs

Overall 465 respondents, LOB=107, EMEA=168

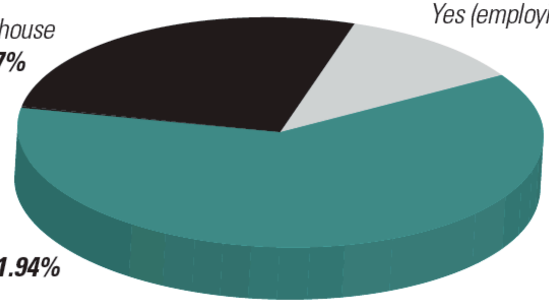
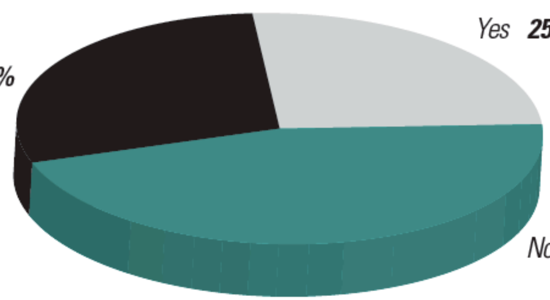


Figure 27a: Organizational Plans to Leverage Social Media Metrics Into Business Processes

Overall 459 respondents, LOB=107, EMEA=167



Brands in SM

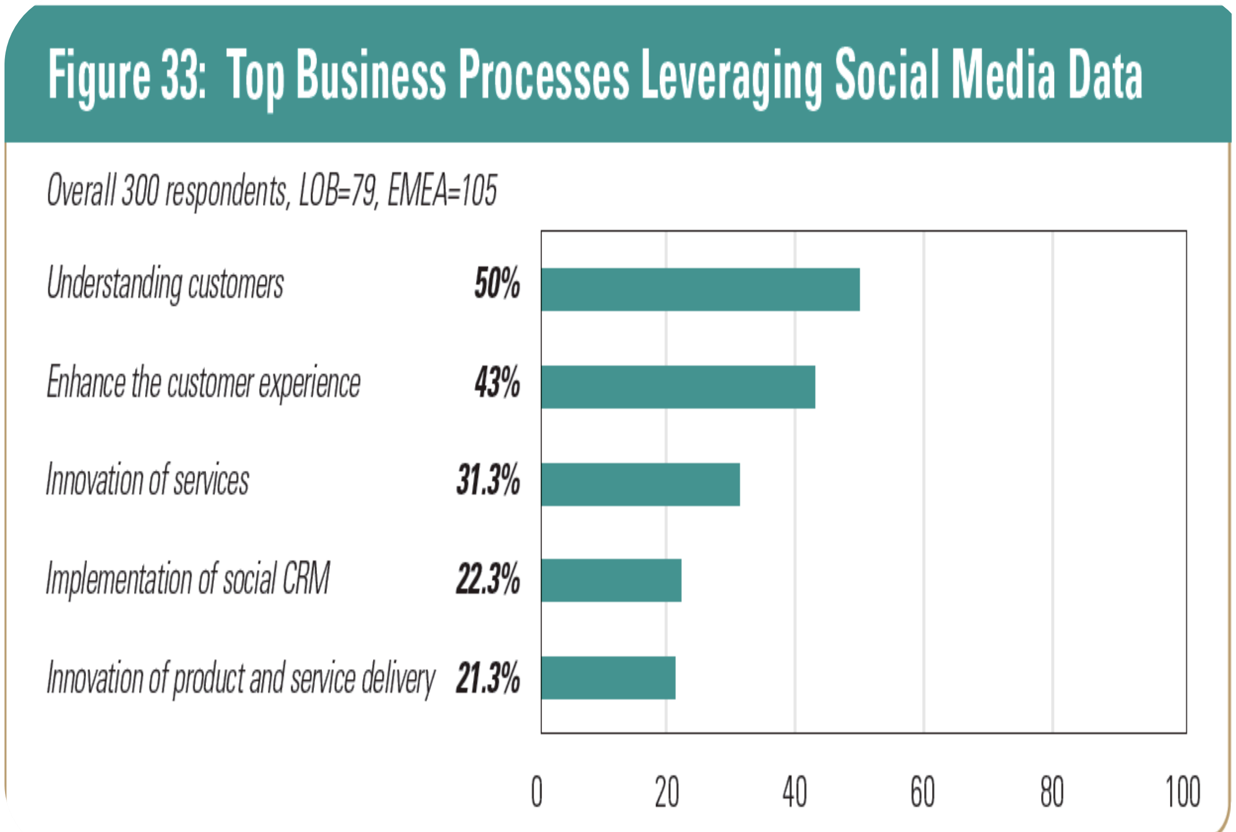


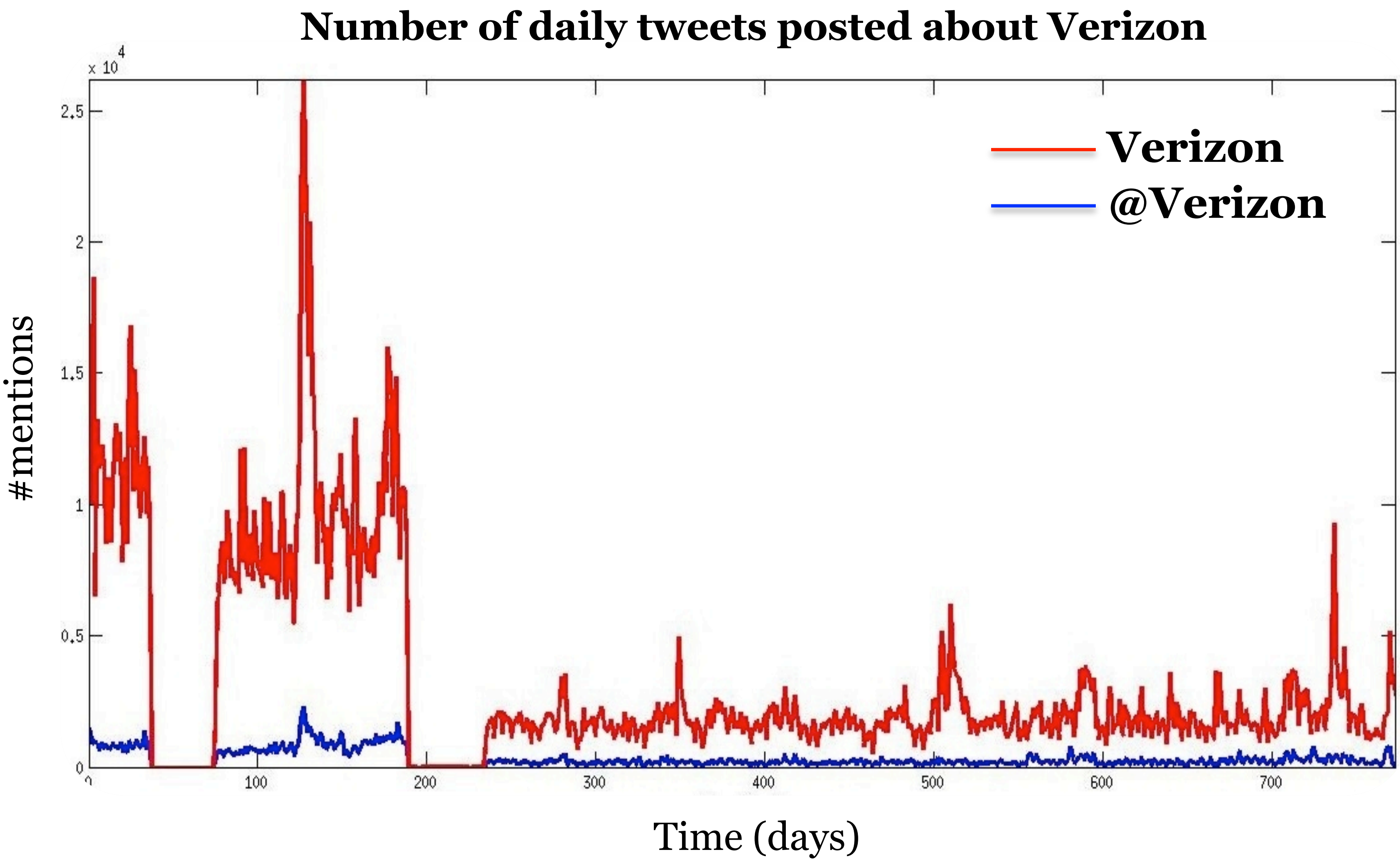
Figure 33: Top Business Processes Leveraging Social Media Data

Large-scale Data

779 K

599 K

Large-scale Data



Challenges

- Slang and words not in dictionaries

Challenges

- Slang and words not in dictionaries
  - **Tomorrow**: 2m, 2ma, 2mar, 2mara, 2maro, 2marrow, 2mor, 2mora, 2moro, 2morow, 2morr, 2morro, 2morrow, 2moz, 2mr, 2mro, 2mrrw, 2mrw, 2mw, tmmrw, tmo, tmoro, tmorrow, tmoz, tmr, tmro, tmrow, tmrrow, tmrrw, tmrw, tmw, tomaro, tomarow, tomarro, tomarrow, tomm, tommarow, tommarrow, tommoro, tommorow, tommorrow, tommorw, tommrow, tomo, tomolo, tomoro, tomorow, tomorro!

Challenges

- Slang and words not in dictionaries
  - Tomorrow: 2m, 2ma, 2mar, 2mara, 2maro, 2marrow, 2mor, 2mora, 2moro, 2morow, 2morr, etc.
- Tweets are short, context-less, and very noisy
  - May not carry desired signals for info extraction

Challenges

- Slang and words not in dictionaries
  - Tomorrow: 2m, 2ma, 2mar, 2mara, 2maro, 2marrow, 2mor, 2mora, 2moro, 2morow, 2morr, etc.

- Tweets are short, context-less, and very noisy
  - May not carry desired signals for info extraction

- Tweets are of streaming type
  - Incoming data may represent new information.

Challenges

- Slang and words not in dictionaries
  - Tomorrow: 2m, 2ma, 2mar, 2mara, 2maro, 2marrow, 2mor, 2mora, 2moro, 2morow, 2morr, etc.
- Tweets are short, context-less, and very noisy
  - May not carry desired signals for info extraction
- Tweets are of streaming type
  - Incoming data may represent new information.
- Very dynamic community structure
  - Follower/Followee relations.

Q1. Churn Prediction

- Churn happens when a customer leaves a brand or stop using its service.
- Churn rate indicates
  - Customer response
  - Average time an individual remains a customer.

Task
Given a **tweet** about a **target brand**, determine if the user intends to leave the brand!

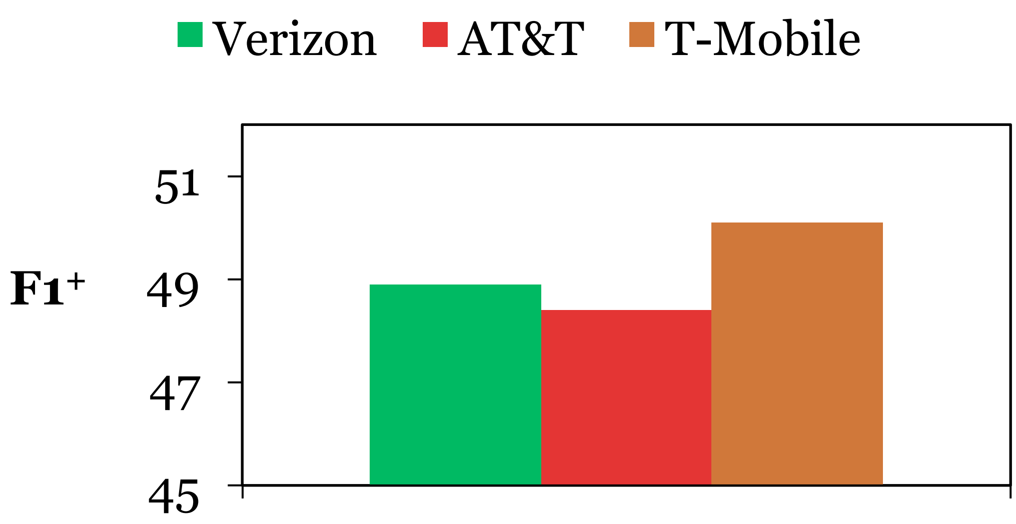
Q1. Churn Prediction

Cant wait to leave [Brand-1] for [Brand-2]! One more bill!!
My days with [Brand-1] are numbered.

I will change carriers as soon as contract is up.
This loyal customer will be gone #awfulcustomerservice.

Isn't It Easy?

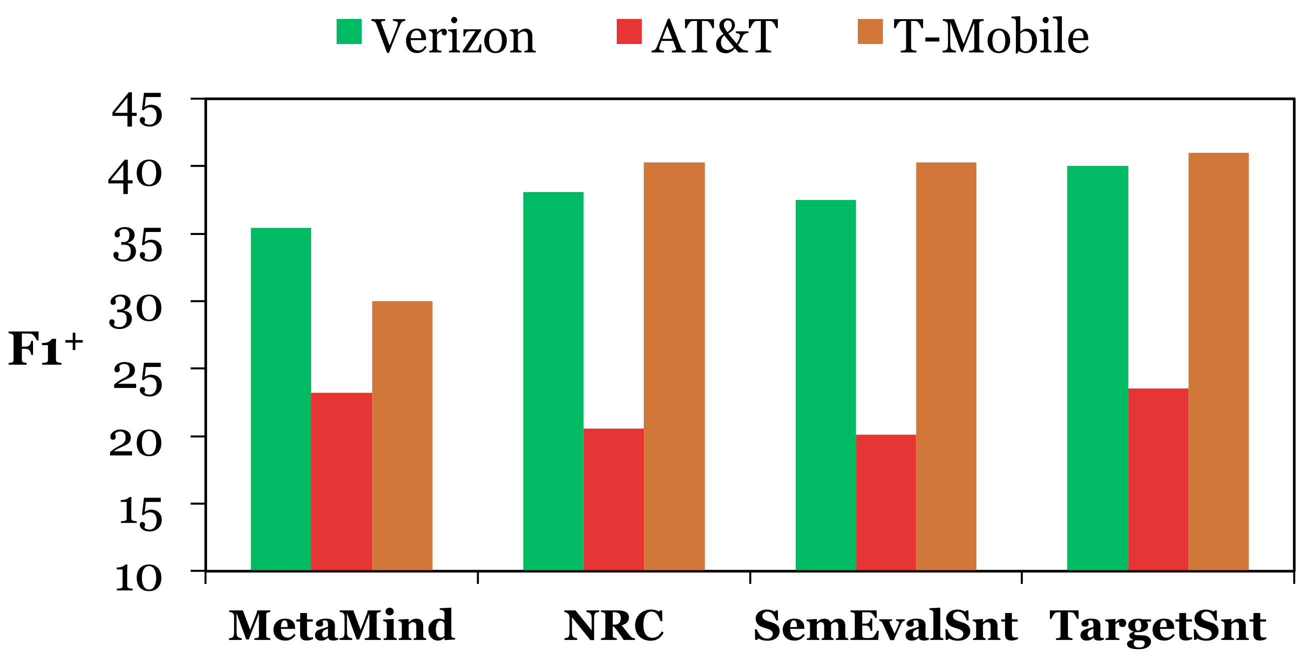
- Use list of churny keywords* to classify tweets as churny or non-churny.



* {leave, leaving, switch, switching, numbered, cancel, canceling, discontinue, give up, call off, through with, get rid, end contract, change to, changing, . . .}

Isn’t it Sentiment Classification?

- Negative tweets → churny
- Positive tweets → non-churny!



- **Positive & Churny**: hate that I might end up leaving [Brand] cuss they are the best company ever
- **Negative & Non-churny**: [Brand]’s cell coverage still sucks

Challenges

1. Target-dependent task: tweets comparing several brands!

I am leaving [BRAND-1] for [BRAND-2]

Challenges

2. Simple language constituents

switch to [Brand]

switch from [Brand]

Challenges

3. Negation effect!

[Brand] is awesome, I'll never leave them

Challenges

4. Churny keywords may not convey churn!

I need a little [Brand]'s #help b4 leaving the states

Challenges

5. Subtle ways in using language as in

debating if I should stay with [Brand]

Important Features

- Demographic

Description
Activity ratio: average No. of posts about brand/competitors per day ratio of active days about brand/competitor average time gap between posts about brand/competitor ratio of urls in posts about brand/competitor average No. of words in post about brand/competitor
Average of friends activity ratios
followers and friends
If user has bio information
If bio contains URL

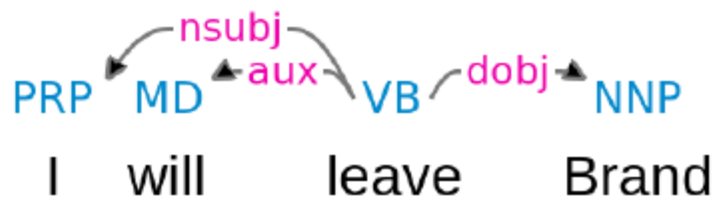
- Content

Description
Unigrams / Bigrams
Neighboring words of brand/competitors names
Syntactic and Comparative marker features
Sentiment features
Tense of tweet
News indicator features

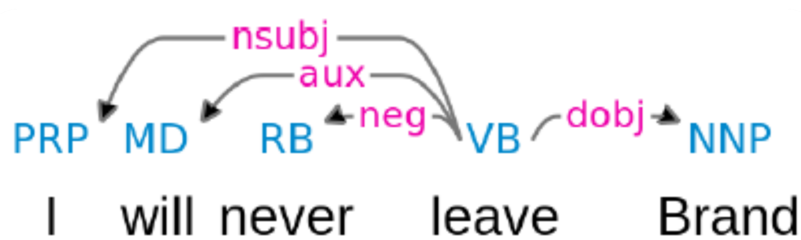
- Context

Description
Content features of user/friends/brand/competitors posts in thread (as defined in Table 4)
posts from user/friends/brand/competitors in thread
posts in thread
Reciprocity between user and brand/competitors posts

Tackling Negation



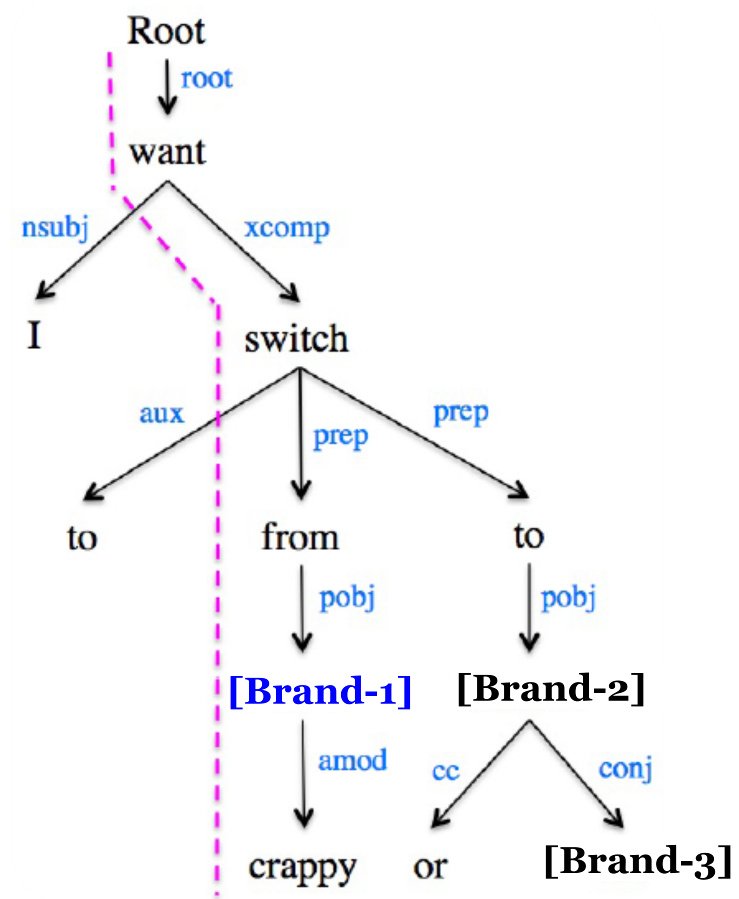
(a) syntactic features: {*dobj-leave-Brand*, *nsubj-leave-i*, *aux-leave-will*}.



(b) syntactic features with negation effect: {*Neg-dobj-leave-Brand*, *Neg-nsubj-leave-i*, *Neg-aux-leave-will*}

Dependency Path

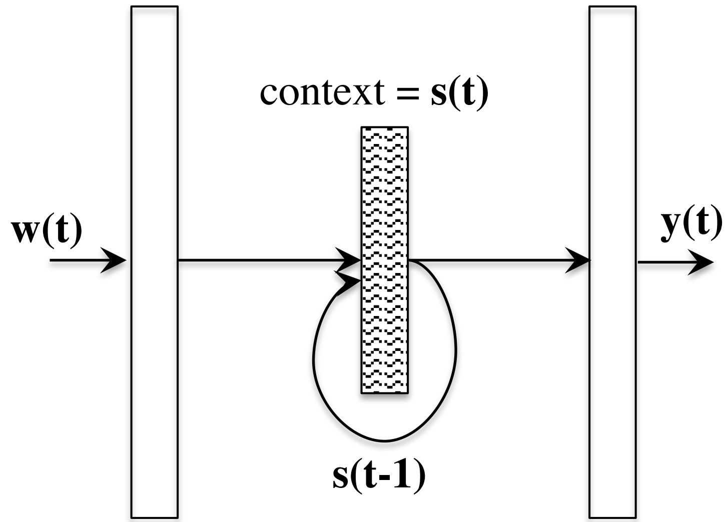
- The sub-tree that covers the path from the root of the tree to the target brand node and all its children.
- Extracts key content for churn prediction



I want to switch from crappy [Brand-1] to [Brand-2] or [Brand-3]

Tweet Representation

- Content representation using Recurrent Neural Network



Basic RNN for tweet Representation

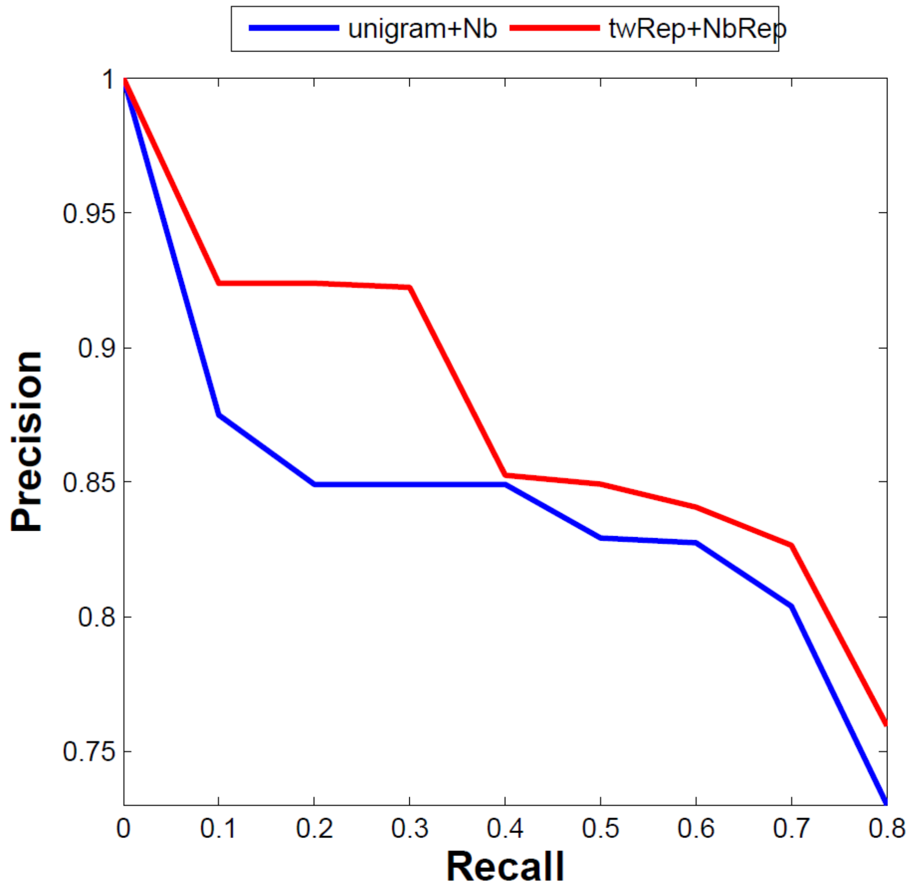
Evaluation- Detection

- Macro-Average Performance over Verizon, T-Mobile, AT&T datasets

	BOW			RNN		
	BOW Features	**hinge**	**logistic**	**hinge**	**logistic**	**RNN Features**
(1)	unigram	65.30	64.30	62.97	63.73	tRep
				59.97	61.37	wRep
				66.13*	**66.20***	twRep
(2)	unigram+Nb	73.63	72.17	71.07	**73.90***	twRep+NbRep
(3)	unigram+Dep	72.40	71.80	**75.66***	75.43*	twRep+DepRep
(4)	unigram+Cntx	74.27	73.20	**75.47***	75.03*	twRep+CntxRep
(5)	unigram+Nb+Dep+Cntx	77.03	75.60	76.77	**77.56***	twRep+NbRep+DepRep+CntxRep
(6)	**BOW+RNN:** hinge: **78.30**, logistic: 78.15					

Evaluation- Ranking

- A good model should rank the most churny tweets higher in its ranking list of churny tweets.



Summary

- Effective techniques for churn prediction in social media.
- Demographic, content, and context features are important.
  - Dependency path & context features
- Churn prediction is not sentiment analysis.