Anthus News

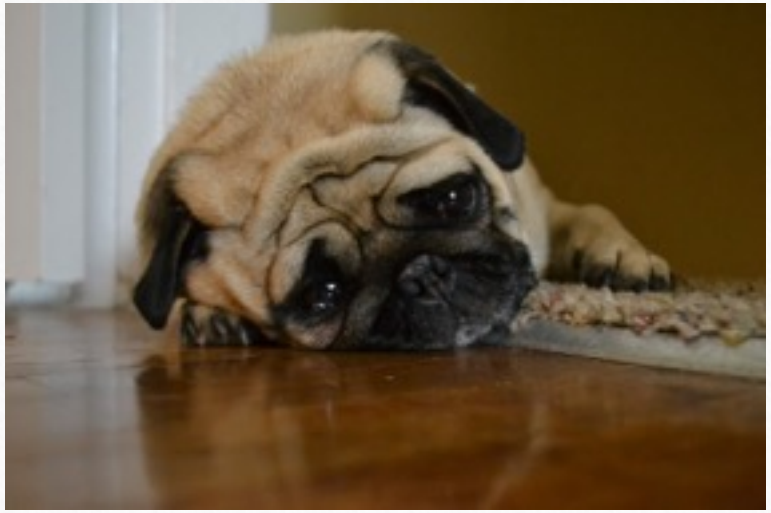
A Personalised News Feed

Robin Businessport:

"60% of the news on this page is not of interest to me … *I want a news feed that reflects my interests …but I don't want to have to fill in forms to tell them what I like. Make it easy to sign in and hassle-free*"

Bob Stats:

"I'd like the app to show me what it thinks my interests are with some analytics."



Sue Showme:

"Tell me what you think my interests are, but let me change them too. I'd like the system to react to what I'm reading and give me more of the same."

Personalising News

To provide a personalised news feed we need to know the user's interests.

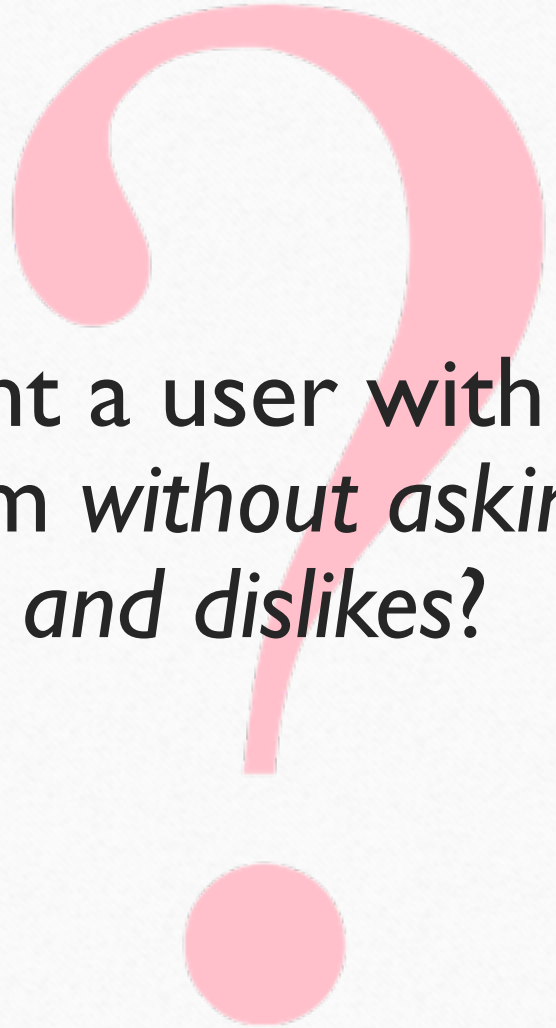
However:

- Users generally dislike explicitly expressing their interests [1]
- Studies have shown that explicit interests do not consistently match behaviour [2]

1. Doychev, Lawlor, et al. "An Analysis of Recommender Algorithms for Online News." CLEF, 2014.
2. Lavie, Talia, et al. "User attitudes towards news content personalization." *International journal of human-computer studies* 68.8 (2010): 483-495

Target User

- Gets their news from *online sources*
- Cares about the *relevance* of the news they read



How to present a user with news articles relevant to them *without asking for their likes and dislikes*?

Our inspiration

Inferring User Interests in the Twitter Social Network

ABSTRACT

We propose a novel mechanism to infer topics of interest of individual users in the Twitter social network. We observe that in Twitter, a user generally follows *experts* on various topics of her interest in order to acquire information on those topics. We use a methodology based on social annotations

We use a methodology based on social annotations to first deduce the topical expertise of popular Twitter users, and then transitively infer the interests of the users who follow them.

models (Labeled LDA) on tweets. Based upon th posed methodology, we build a system **Who Likes Wh** which can infer the interests of *millions of Twitter user* our knowledge, this is the first system that can infer s for Twitter users at such scale. Hence, this syste be particularly beneficial in developing personalize mender services over the Twitter platform.

Bhattacharya, Parantapa/Muhammad Zafar/Niloy Ganguly/Saptarshi Ghosh/Krishna Gummadi "Inferring user interests in the Twitter social network" (2014): 357–360. doi:10.1145/2645710.2645765

DEMO

http://csi6220-2-vm4.ucd.ie

The System

Our inspiration

Inferring User Interests in the Twitter Social Network

ABSTRACT

We propose a novel mechanism to infer topics of interest of individual users in the Twitter social network. We observe that in Twitter, a user generally follows *experts* on various topics of her interest in order to acquire information on those topics. We use a methodology based on social annotations

We use a methodology based on social annotations to first deduce the topical expertise of popular Twitter users, and then transitively infer the interests of the users who follow them.

models (Labeled LDA) on tweets. Based upon th posed methodology, we build a system **Who Likes Wh** which can infer the interests of *millions of Twitter user* our knowledge, this is the first system that can infer s for Twitter users at such scale. Hence, this syste be particularly beneficial in developing personalize mender services over the Twitter platform.

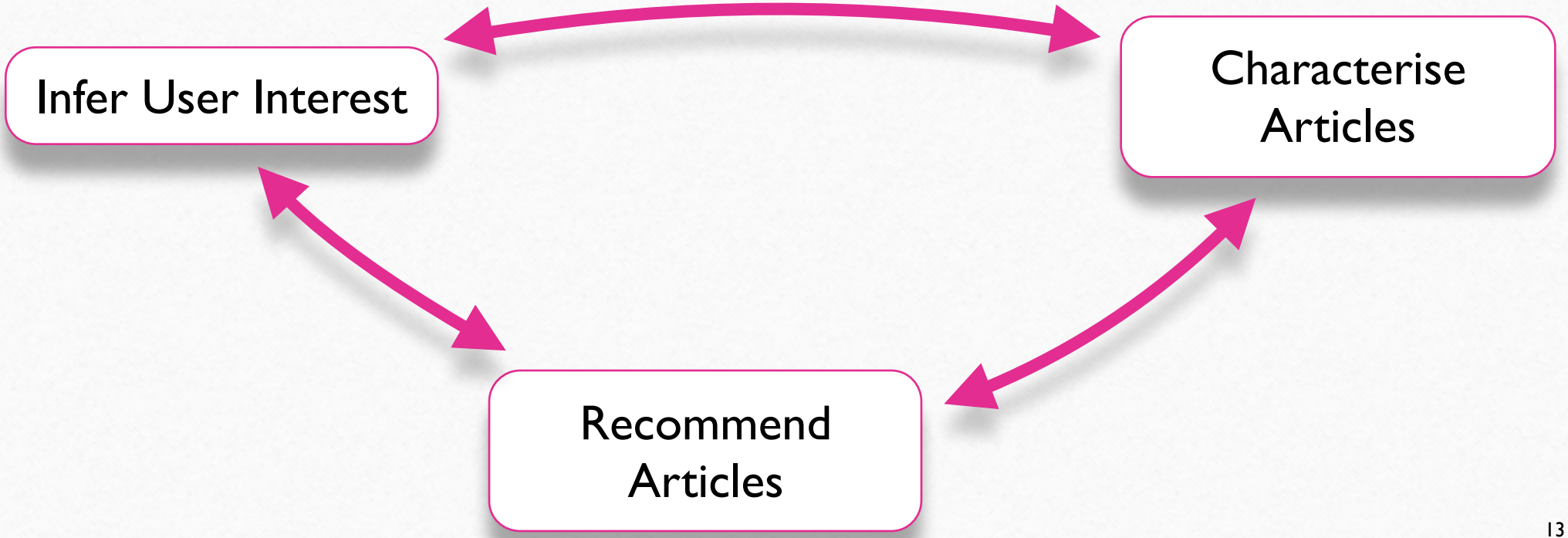
Bhattacharya, Parantapa/Muhammad Zafar/Niloy Ganguly/Saptarshi Ghosh/Krishna Gummadi "Inferring user interests in the Twitter social network" (2014): 357–360. doi:10.1145/2645710.2645765

Our modifications to their method

Their version	Our version
Take whole Twitter database from 2009 and add in every list added since. Use HITS algorithm to identify hubs.	Load every list on which the top 1000 Twitter users appear (120,000 lists, 9.9 million unique users ~5% of Twitter active users)
Generate 1800 inferred topics (from list data)	Select 15 topics
Identify v_t (>10 mentions of t on v's lists)	Identify v as those on >10 lists and associate them with multiple t using the text from their lists (v_t is an array)
Consider user interested in t when they follow 3 or more v_t	Consider user interested in t when they follow 1or more v_t

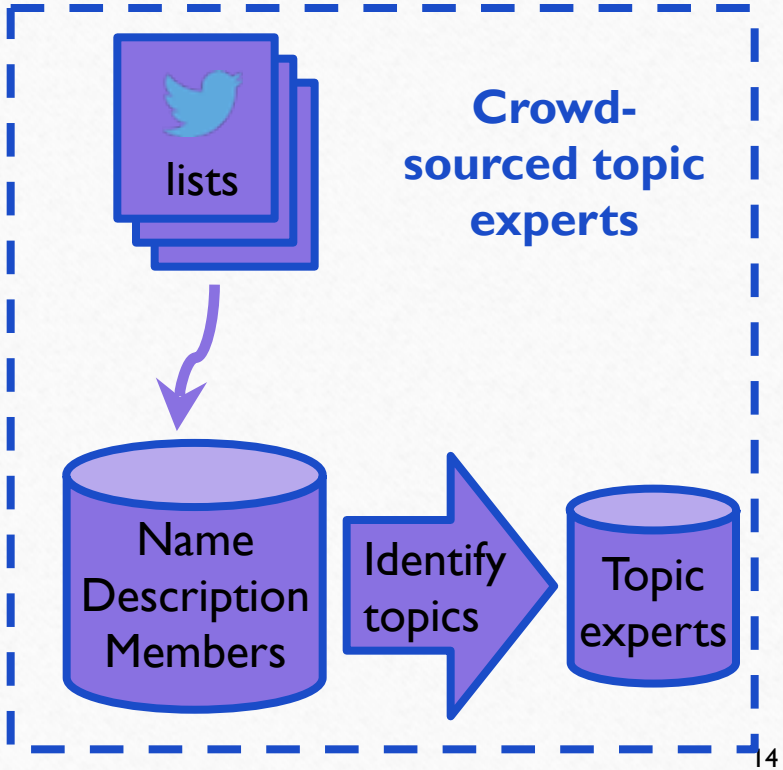
t: topic v: expert v_t : expert in topic t

Technical Challenges

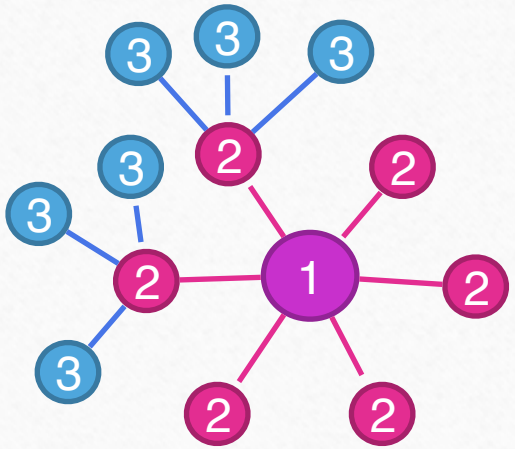


Offline: Inferring topics and experts

- Parse names, descriptions of Lists
  - Identify top terms by frequency across all lists
  - Select 16 of these terms as topics
- Identify experts
  - 10 mentions of topics on a user's Lists



Loading Twitter Lists

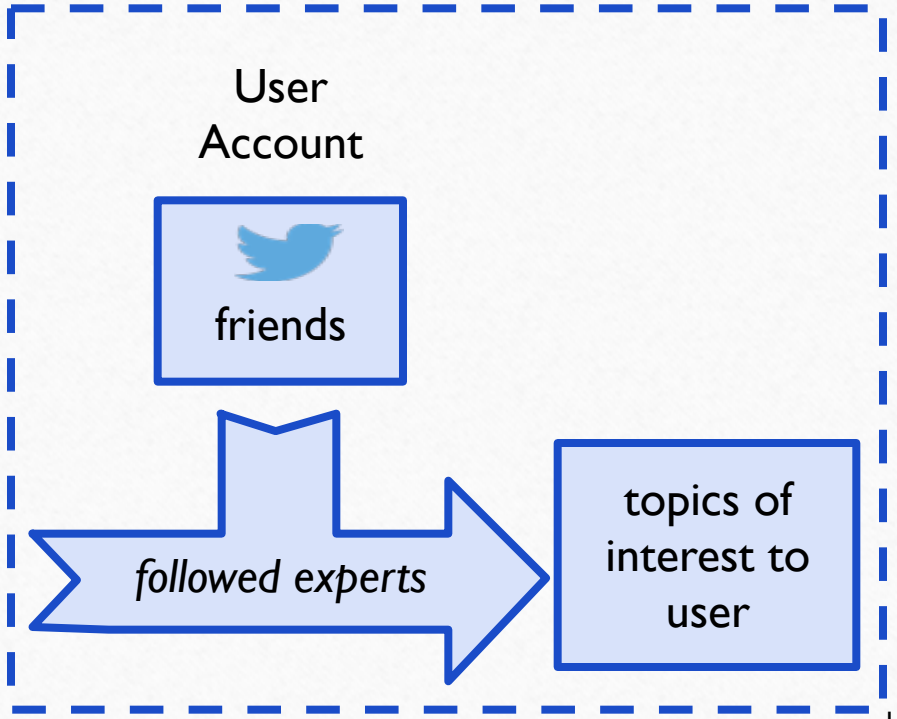


1. WFS algorithm based on top 1000 users
2. Load lists for Twitter users sequentially by id

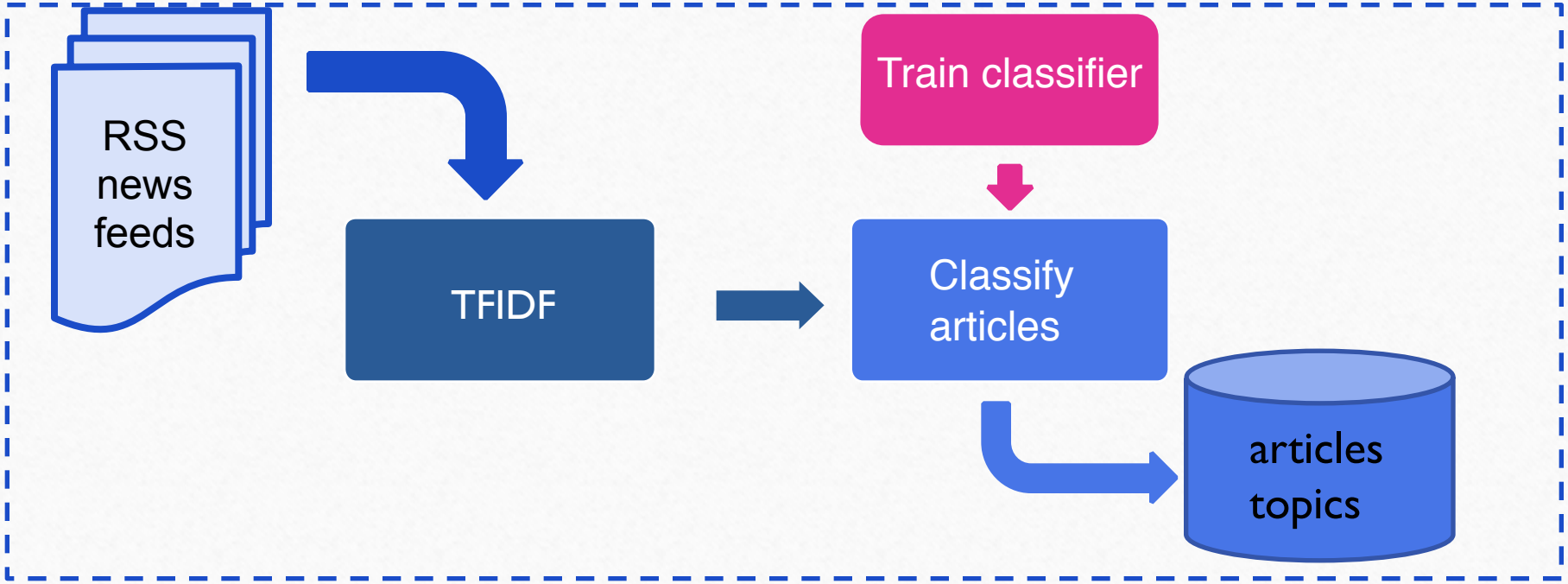
Online: Inferring user interests

Compute Borda count of topics in a user's friend list

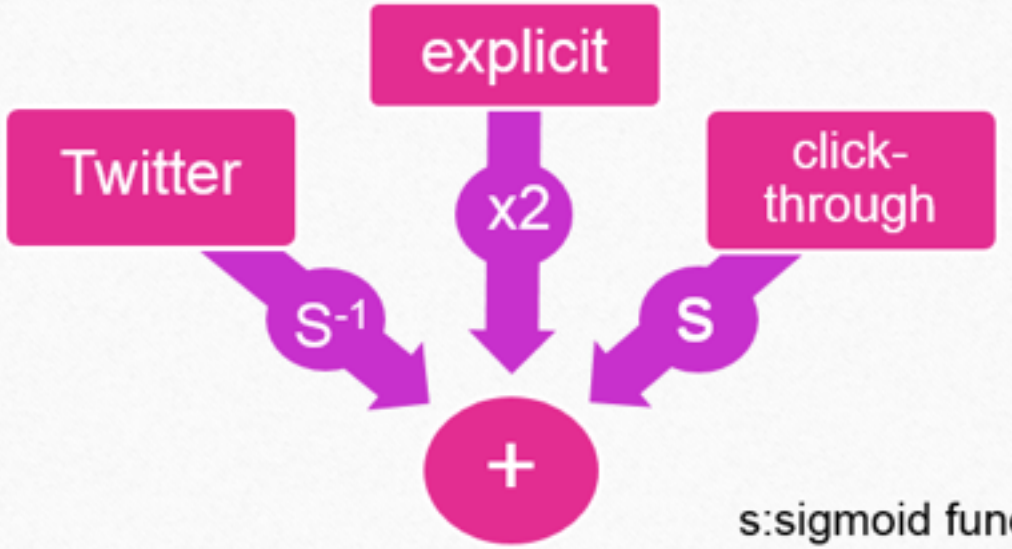
- Rank topics per expert by frequency
- Topic 1 - 5 points
- Topic 2 - 4 points etc…
- Sum over entire friend list



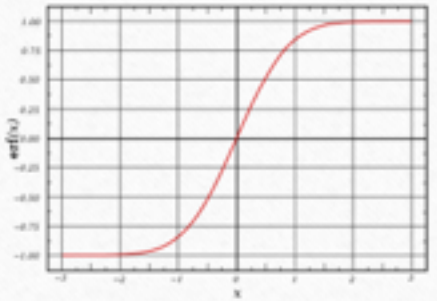
Characterising news articles



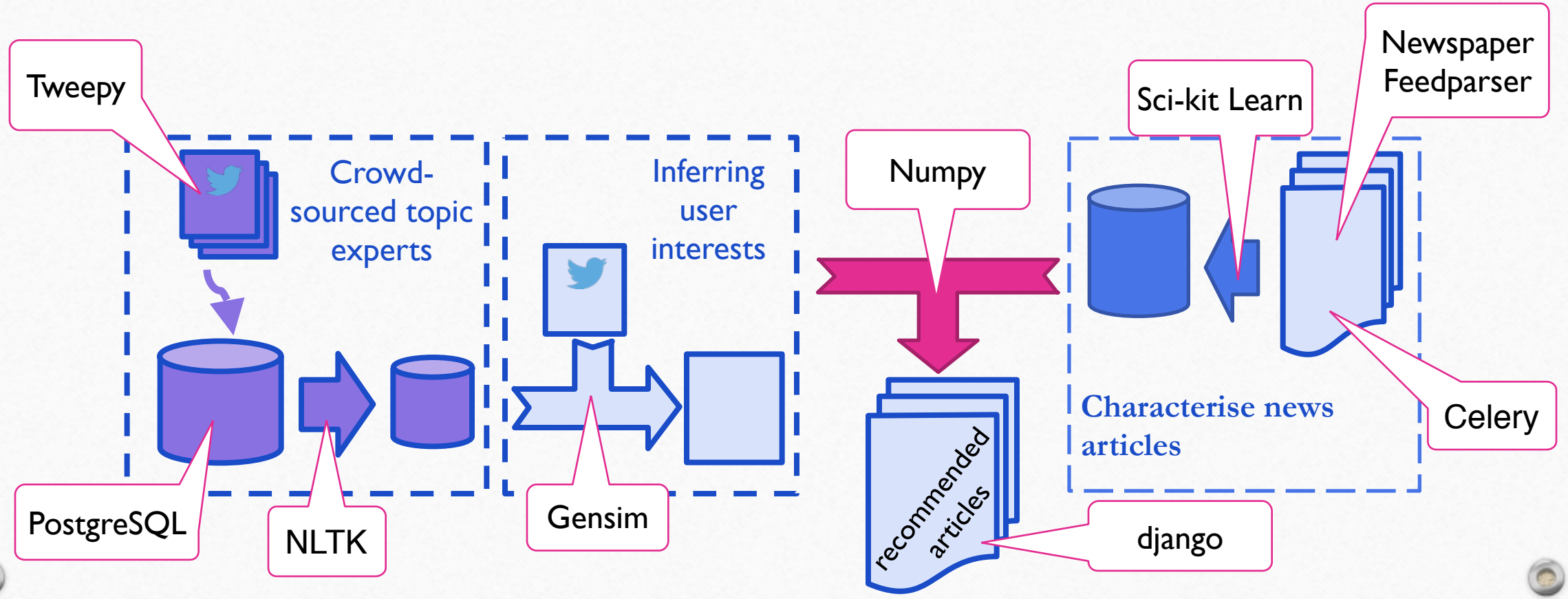
Recommend Articles: weighting function



s:sigmoid function



System Architecture



Design Objectives

- Provide a personalised experience for users
- Generate user interests implicitly, with minimal input from users
- Make sign-up as streamlined as possible
- Allow users to customise their profiles if they wish to do so
- Give users feedback on the statistics we have collected

Design Considerations

- Fast (asynchronous processing where appropriate)
- Minimalistic design
- Frequent trawl for new articles

Business Proposition

Customer Segments

General Users

Free service

Create value by providing personalised service

Large news outlets

Free, mutually beneficial

Generate traffic

Small News Outlets

Inclusion fees

Generate traffic, raise online profile

Profit Model

- Sponsored articles
  - Pay-per-click
- Inclusion fees
  - Annual fee
- Advertising
  - Pay-per-click

Early Adopters - €29
General - €99

Several price points depending on traffic

Processes

- Design & Development
- Promotion & Awareness
- Monetisation
- Maintenance

Facebook & Twitter

Use Shopify for e-commerce

Review

Evaluation - UX

- Survey based on Nielsen's heuristics
  - Invited Twitter users to use system
- Beta Testers provided feedback
  - User interface
  - Bugs
  - Functionality

Article Classifier: manual review

70%

Art	100%		Business	40%	sport
Entertainment	70%	sport	Celebrity	50%	sport, tech
Music	55%	sport	Tech	70%	design
Education	75%		Film	70%	sport, celebrity
Science	65%		Sport	100%	
Design	40%	sport	Food	90%	
Politics	75%		Travel	50%	range of topics
Fashion	100%		Health	55%	business

Comparison with Wholikeswhat

88%

		wholikeswhat	anthus	✓
ART	theellenshow	Celebrity, entertainment	Entertainment,art, music	✓
FOOD	Bourdain	Travel,food	Food,entertainment, travel	✓
MUSIC	Justin Bieber	music	music,art	✓
EDUCATION	edutopia	education	tech, education	✓
SCIENCE	natgeo	science	science, travel, tech	✓
DESIGN	mashable	media,tech,design	tech,design	

Problem - Tech Bias

- Many users "experts" in technology
  - Order of list loading
  - Inherent to Twitter
  - "Hidden" feature

- Solution
  - More lists
  - Sub-categories

Problem - Classification Errors

- Some articles well-classified

Simpsons-Inspired Heavy Metal Band “Okilly Dokilly” Consists Of 5 Neds

What's the most unlikely inspiration for a metal band? Ned Flanders, of course, and that's why Okilly Dokilly is the only Flanders-themed band out there. Created by Head Ned and Bled Ned (probably not real names), it has three other members: Red Ned, Thread Ned and Stead Ned. And they play metal, with lyrics that are basically all Ned Flanders quotes. For example, “All That Is Left”, is a song dedicated to the Leftorium store.

Aug. 14, 2015, 3:34 p.m.

♫ music

Problem - Classification Errors

- Some not so well

White Sox third baseman Tyler Saladino had a memorable defensive inning

As the old baseball joke goes, Chicago White Sox third baseman Tyler Saladino tied a long-standing record by being involved in all three outs in the second inning on Saturday. The joke will always work too, because it's a record that has been tied thousands of times over the years, but will never be broken. That much is guaranteed. [Yahoo Sports Fantasy Football: Sign up and join a league today!] The reason it's worth mentioning this time though has nothing to do with jokes or records, it's the manner in which Saladino recorded those outs. Each play was different from the others. Each play had a certain level of difficultly attached. And, above all else, each out was completed in impressive fashion.

Aug. 16, 2015, 5:05 p.m.

♫ music

Problem - Classification Errors

- Small number of topics (too broad, many overlaps)
- Small training set (20-30 articles per topic)

- Solution
  - Increase training set size
  - More topics
  - Introduce topic hierarchy

Improvements

- Improve classifier
- Aggregate more List data
- Recommend topics/articles to users
- Recommend Twitter accounts

Reflections

Success?

- Twitter Lists
- Analytics
- User Profiles
  - Implicit
  - Explicit
  - Click-throughs

- Topics
- Classification

Challenges

- Getting started!
- Github…
- Working with limited resources
- Twitter rate limits

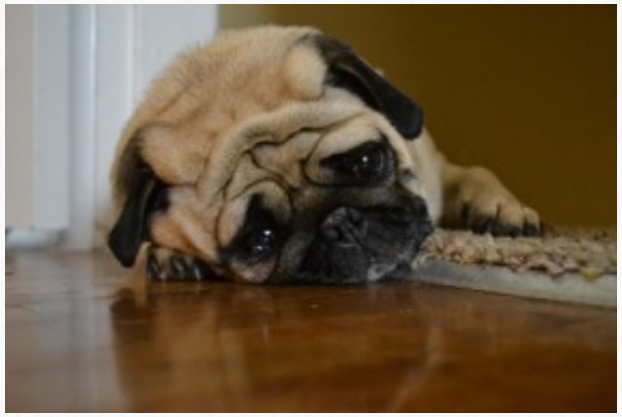
Lessons Learned

- Skills
  - Python
  - Database
  - NLP
- Agile methods
- Research skills

- Importance of communication
- Coordination/ delegation

Remember Bob and Sue?

"I'd like the app to show me what it thinks my interests are with some analytics."



"Tell me what you think my interests are, but let me change them too. I'd like the system to react to what I'm reading and give me more of the same."

ANTHUS NEWS

WELCOME, JIMMMMMMY | MY NEWS | ANALYTICS | MY PROFILE | CONTACT | LOGOUT

My Profile

My Topics

Top interests from Twitter

Questions?