Crossing Media Streams with Sentiment: Domain Adaptation in Blogs, Reviews, and Twitter

Yelena Mejova, Padmini Srinivasan
University of Iowa

June 6, 2012 ICWSM


digg SHARE
SHARE
SHARE
SHARE
SHARE
SHARE
SHARE
You Tube SHARE
SHARE
SHARE
SHARE
SHARE
SHARE
SHARE
in SHARE
SHARE
SHARE
SHARE
SHARE
SHARE
SHARE
PIANO
Bē SHARE
SHARE
SHARE
MV SHARE
SHARE

ICWSM chatter

backchan.nl

Social grooming and task completion views of social networks in the same day. Where is the balance?

"Context collapse", when social circles collide. Worse for high "self-monitors" (Snyder, 1974), a.k.a. social chameleons.

There is no delete on the Internet! Your information still exists in a cache or an archive...

Love this Q from @munmun10: How would review ratings, particularly of deceptive reviews differ for systems driven by review incentive? E.g. eBay

Quite interesting talk by Song Feng. Recently we Japanese encountered a similar problem in a kind of Michelin guide called Tabelog, which led to losing the trust of many users. From this viewpoint, we hope this work well organized :)

twitter

Simon Lindgren @simon_lindgren 1m
Oversharing on the one hand. Privacy concerns on the other. What is the motivation behind not sharing? #icwsm
Expand

Dr Sina Samangooei @sinjax 1m
Reason I never use geolocated services on social networks. When it matters, battery life often matters more :-) #icwsm
Expand

Simon Lindgren @simon_lindgren 2m
don't disturb my circles! on boundary preservation and location sharing in social media (Xinru Page ics.uci.edu/~xpage/ at #icwsm)
Expand

Marie Boran @pixievondust 3m
Great talk by @nicole_ellison on Facebook, privacy and social capital. #icwsm
Expand

Armando Alves @armandoalves 4m
Challenge: How can we design social web services to maximize social capital while maintaining privacy #icwsm
Expand

UX Feeder @UXfeeder 5m
Delicious: Proceedings of the Sixth International Conference on Weblogs and Social Media: bit.ly/LkFa7A [Research]
Expand

google+

Amit Sheth 12:25 PM - Public
+Andrew Tomkins gave an excellent keynote at #icwsm - this slides shows all the exciting metadata waiting for deep #socialmedia analysis: http://t.co/sikjSD1B [This seems more metadata than Twitter-- so systems like Twitris: http://twitris.knoesis.org can perhaps even deeper!]

http://t.co/sikjSD1B »

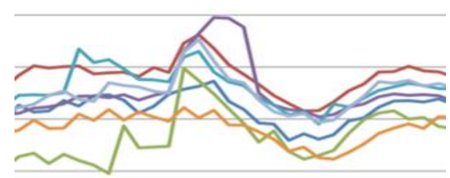
Marco Guerini 12:10 PM - Mobile - Public
"Distributional footprints of deceptive product review"... Impressive work, I enjoyed it! (This is not a fake review, of course!) #icwsm

+1

Questions

- Does sentiment on a topic differ between social media sources?
- To what extent can an SA classifier developed on one social medium can be transferred to another?

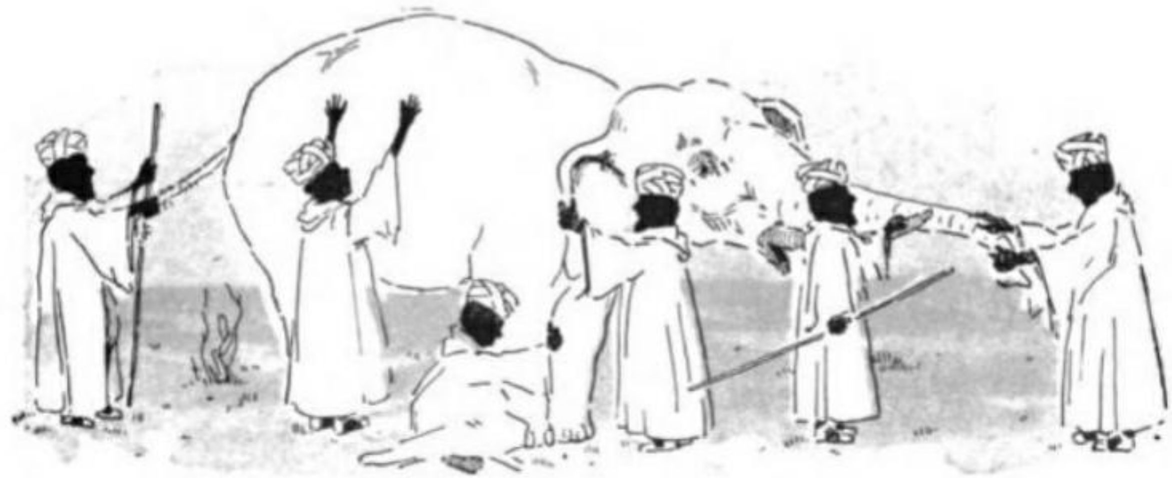
source 1



source 2

Why?

- Is one stream indicative of large picture? Unlikely



- Can we utilize information across streams?

Contributions

- Create a multi-source, multi-topic dataset
- Compare POS, NEG, MIX, NEU sentiments
- Cross-source adaptation experiments:
  - Single-source
  - Multiple-source
    - Mixed models
    - Voting models

Gathering Data

- Streams: Twitter, Blogs, Reviews

* Most popular

Topic Categories	Initial Topic Source
movies	IMDB
musical albums	Amazon
computer games	Metacritic
smart phones	CNet
restaurants	Yelp

For each stream:
    Starting from the top, for each topic:
        If # of returned results < 50 → discard topic, else keep topic
        If # of returned results > 100 → select 100 randomly
Until each stream has min 500 documents retrieved

Gathering Data

Topic Categories	Reviews	Twitter	Blogs
movies	IMDB	Searched using Twitter Search API, discarding tweets fewer than 10 characters in length	Results from a Google Blog Search crawled, content heuristically extracted, min length of 100 chars
musical albums	Amazon		
computer games	Amazon		
smartphones	CNet		
restaurants	Yelp		

tag density – proportion of HTML tags of N consecutive tokens

Gathering Data

Category	#queries		For chosen, # of docs			After cleaning & sampling		
	Tried	Chosen	Twitter	Reviews	Blogs	Twitter	Reviews	Blogs
Movies	19	8	7538	17707	510	770	800	423
Music	56	8	5738	1760	556	740	772	467
Games	7	7	2193	2123	584	495	617	525
Phones	12	5	7479	2146	573	482	500	432
Rest-ts	26	9	881	22731	614	566	900	355
Total	120	37	23829	46467	2837	3053	3589	2202

ambiguity

limiting stream

duplication
(retweeting)

Labeling

- Twitter & Blogs
- Quality control
  - Obvious question
  - Enter first word of last tweet
- Majority out of 3
  - Topicality:
    - *Yes*, *No*, *Can't Tell*
  - Sentiment:
    - *Positive*, *Negative*, *Mixed*, *None*, *Can't Tell*

Category	Total	Topical	Not top.	Other	Pos	Neg	Mix	None	Oth
	Blogs								
Movies	423	184 (44)	196 (46)	43 (10)	87 (47)	12 (7)	29 (16)	46 (25)	10 (5)
Music	462	243 (53)	160 (35)	59 (12)	154 (63)	8 (3)	19 (8)	51 (21)	11 (5)
Games	525	285 (54)	187 (36)	53 (10)	154 (54)	20 (7)	32 (11)	60 (21)	19 (7)
Phones	427	261 (61)	136 (32)	30 (7)	130 (50)	17 (7)	33 (13)	60 (23)	21 (8)
Rest-nts	355	138 (39)	172 (49)	45 (12)	96 (70)	2 (1)	17 (12)	15 (11)	8 (6)
Total	2192	1111 (51)	851 (39)	230 (10)	621 (56)	59 (5)	130 (12)	232 (21)	69 (6)
	Twitter								
Category	Total	Topical	Not top.	Other	Pos	Neg	Mix	None	Oth
Movies	770	612 (80)	126 (16)	32 (4)	182 (30)	41 (7)	16 (3)	319 (52)	54 (9)
Music	740	731 (99)	3 (0)	6 (1)	263 (36)	10 (1)	10 (1)	397 (54)	51 (7)
Games	495	473 (95)	14 (3)	8 (2)	128 (27)	26 (6)	42 (9)	231 (49)	46 (10)
Phones	482	479 (99)	1 (0)	2 (1)	187 (39)	99 (21)	29 (6)	142 (30)	22 (5)
Rest-nts	566	545 (96)	9 (2)	12 (2)	268 (49)	14 (3)	32 (6)	200 (37)	31 (6)
Total	3053	2840 (93)	153 (5)	60 (2)	1028 (36)	190 (7)	129 (5)	1289 (45)	204 (7)
	Reviews								
Category	Total	Topical	Not top.	Other	Pos	Neg	Mix	None	Oth
Movies	800	800 (100)	–	–	612 (77)	91 (11)	97 (12)	–	–
Music	772	772 (100)	–	–	627 (81)	84 (11)	61 (8)	–	–
Games	617	617 (100)	–	–	504 (82)	63 (10)	50 (8)	–	–
Phones	500	500 (100)	–	–	316 (63)	96 (19)	88 (18)	–	–
Rest-nts	900	900 (100)	–	–	715 (78)	70 (8)	115 (13)	–	–
Total	3589	3589 (100)	–	–	2774 (77)	404 (11)	411 (12)	–	–

number of documents (percentage)

all streams favor positive sentiment

Category	Total	Topical	Not top.	Other	Pos	Neg	Mix	None	Oth
Blogs									
Movies	423	184 (44)	196 (46)	43 (10)	87 (47)	12 (7)	29 (16)	46 (25)	10 (5)
Music	462	243 (53)	160 (35)	59 (12)	154 (63)	8 (3)	19 (8)	51 (21)	11 (5)
Games	525	285 (54)	187 (36)	53 (10)	154 (54)	20 (7)	32 (11)	60 (21)	19 (7)
Phones	427	261 (61)	136 (32)	30 (7)	130 (50)	17 (7)	33 (13)	60 (23)	21 (8)
Rest-nts	355	138 (39)	172 (49)	45 (12)	96 (70)	2 (1)	17 (12)	15 (11)	8 (6)
Total	2192	[illegible]	851 (39)	230 (10)	621 (56)	59 (5)	130 (12)	232 (21)	69 (6)
Twitter									
Movies	770	612 (80)	126 (16)	32 (4)	182 (30)	41 (7)	16 (3)	319 (52)	54 (9)
Music	740	731 (99)	3 (0)	6 (1)	263 (36)	10 (1)	10 (1)	397 (54)	51 (7)
Games	495	473 (95)	14 (3)	8 (2)	128 (27)	26 (6)	42 (9)	231 (49)	46 (10)
Phones	482	479 (99)	1 (0)	2 (1)	187 (39)	99 (21)	29 (6)	142 (30)	22 (5)
Rest-nts	566	545 (96)	9 (2)	12 (2)	268 (49)	14 (3)	32 (6)	200 (37)	31 (6)
Total	3053	2840 (93)	153 (5)	60 (2)	1028 (36)	190 (7)	129 (5)	1289 (45)	204 (7)
Reviews									
Movies	800	800 (100)	–	–	612 (77)	91 (11)	97 (12)	–	–
Music	772	772 (100)	–	–	627 (81)	84 (11)	61 (8)	–	–
Games	617	617 (100)	–	–	504 (82)	63 (10)	50 (8)	–	–
Phones	500	500 (100)	–	–	316 (63)	96 (19)	88 (18)	–	–
Rest-nts	900	900 (100)	–	–	715 (78)	70 (8)	115 (13)	–	–
Total	3589	3589 (100)	–	–	2774 (77)	404 (11)	411 (12)	–	–

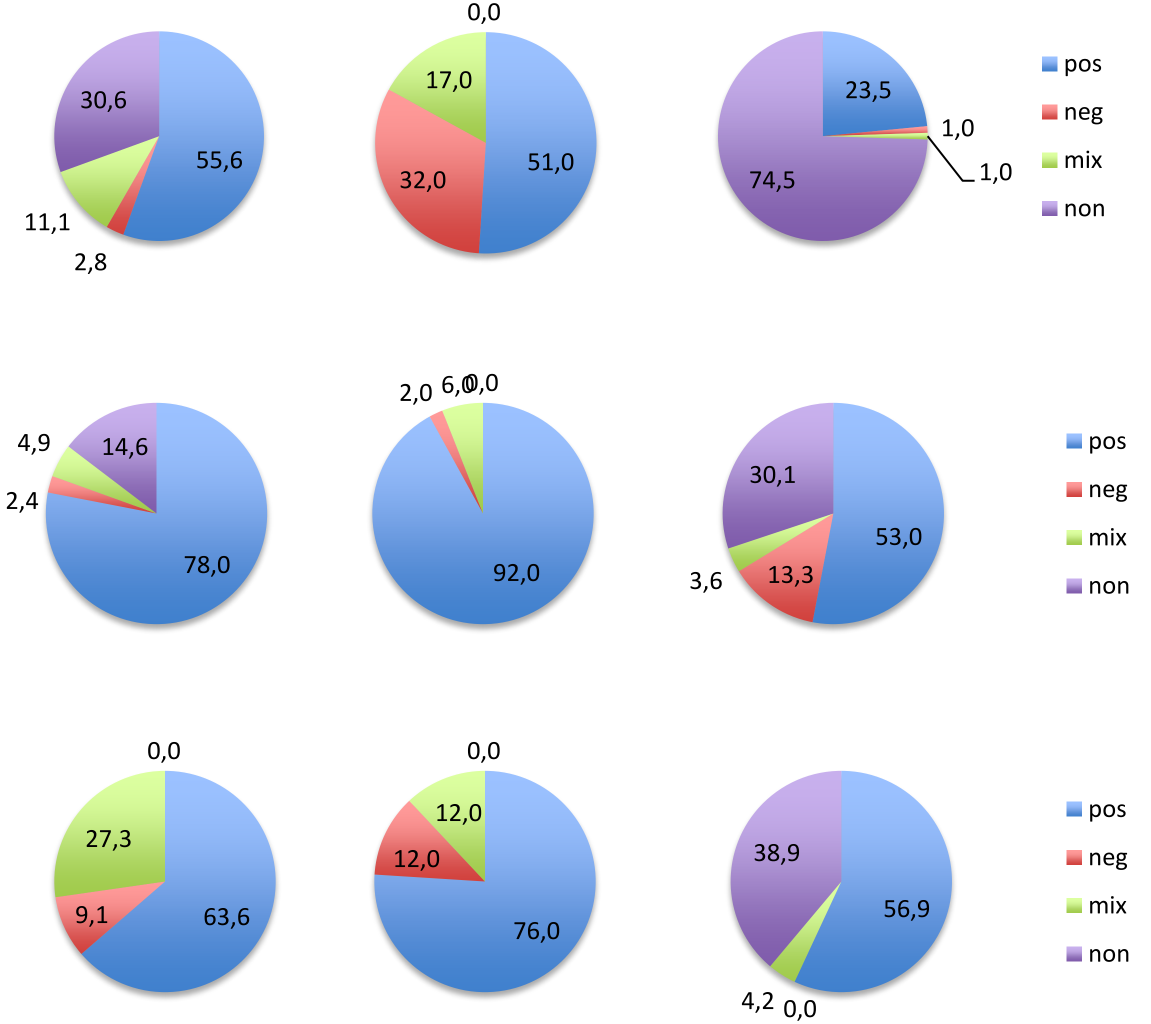
Blogs have more mixed sentiment docs about movies

Phones have unusual negative # of docs in Twitter & Reviews

Twitter contains twice as many docs with no sentiment as blogs

number of documents (percentage)

Blogs | Reviews | Twitter



Sentiment in SM Streams is **Different**

Sentiment Score: % positive - % negative

	all docs			subjective docs		
	B/T	B/R	R/T	B/T	B/R	R/T
movies	30.7	30.1	40.7	28.3	25.1	17.3
music	26.2	24.3	38.2	19.7	27.7	20.8
games	27.9	22.6	50.5	22.9	17.2	25.9
phones	24.7	19.3	25.1	35.7	30.1	20.0
rests	19.9	11.6	23.0	24.8	14.7	13.5
all average	25.8	21.5	35.6	25.6	22.3	19.1
all STD	13.2	16.3	19.9	17.0	15.0	16.8

average difference between sentiment score for a topic

There are substantial differences between sentiments in these streams

Cross-Stream Adaptation

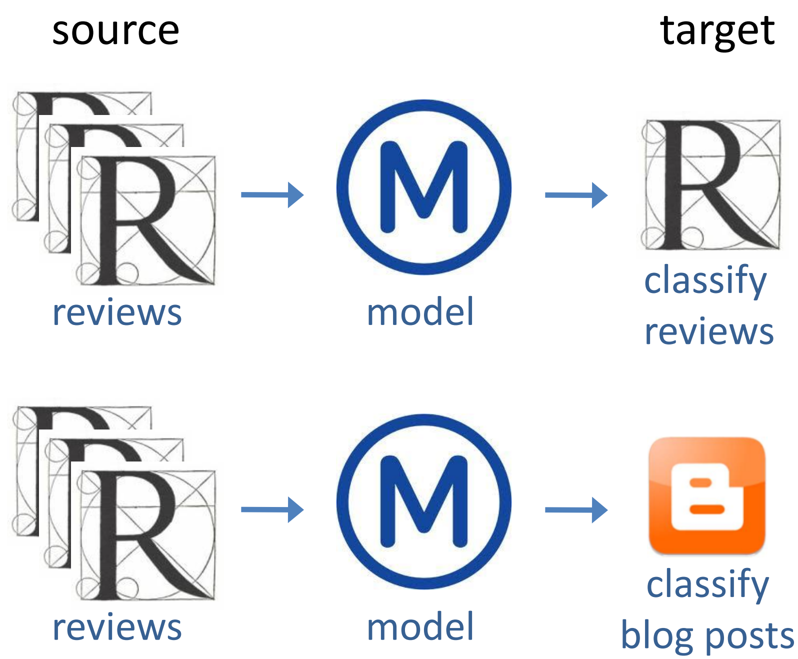
- Lingpipe logistic regression classifier
- Two tasks: detect NEG, detect POS
- 1,2,3-grams

Standard:

(native model)

Adaptation:

(foreign model)



			Accuracy			Target F-score		
Task	Categ.	Target	Blogs	Reviews	Twitter	Blogs	Reviews	Twitter
POS	Games	Blogs	0.631	**0.652†**	0.543	0.763	**0.824†**	0.627
		Reviews	0.788	**0.897**	0.865	0.880	**0.948**	0.927
		Twitter	0.528	0.365	**0.709**	0.520*	**0.646***	0.608*
	Movies	Blogs	**0.655**	0.638*	0.516*	0.783	**0.790***	0.568
		Reviews	0.731	**0.883**	0.759*	0.844	**0.941**	0.861*
		Twitter	0.534	0.367	**0.705**	0.439	**0.612***	0.581
	Music	Blogs	**0.762**	0.708	0.754*	**0.857**	0.848*	0.844*
		Reviews	0.856*	**0.900**	0.878	0.923*	**0.949**	0.937
		Twitter	0.670	0.380	**0.787**	0.558*	0.668*	**0.714**
	Phones	Blogs	**0.635**	0.608*	0.414	0.748	**0.792***	0.410
		Reviews	0.769	**0.815**	0.279	0.869	**0.905**	0.465
		Twitter	0.570*	0.513	**0.610**	0.531*	**0.637†**	0.561
	Rest-nts	Blogs	0.800	**0.814***	**0.814***	0.893	**0.905***	0.896*
		Reviews	0.882*	**0.923**	0.917*	0.938*	**0.961**	0.958*
		Twitter	0.548	0.539	**0.696**	0.651*	**0.753***	0.741
NEG	Games	Blogs	**0.815**	0.794*	0.805*	**0.302**	0.140*	0.126*
		Reviews	0.783	**0.814**	0.809*	0.103	**0.275**	0.037
		Twitter	0.721	0.838*	**0.859**	0.181	0.119	**0.383**
	Movies	Blogs	0.777	0.672	**0.783***	0.096	**0.240†**	0.179*
		Reviews	0.736	**0.802**	0.761*	0.192*	**0.483**	0.236*
		Twitter	0.844	0.800	**0.905**	0.282*	0.185	**0.356**
	Music	Blogs	**0.888**	0.810	0.884*	**0.185**	0.088*	0.000*
		Reviews	0.796	**0.828**	0.811	0.148	**0.435**	0.000
		Twitter	0.953	0.775	**0.978**	0.000*	0.152*	**0.333**
	Phones	Blogs	0.775	0.689	**0.814†**	0.038	**0.250***	0.197*
		Reviews	0.616	**0.698**	0.622	0.141	**0.513**	0.294
		Twitter	0.637*	0.578	**0.704**	0.226*	**0.440***	0.288
	Rest-nts	Blogs	**0.866**	0.837*	0.859*	**0.222**	0.000*	0.000*
		Reviews	0.764*	**0.802**	0.797*	0.186*	**0.354**	0.116*
		Twitter	0.863	0.874*	**0.920**	0.000	0.129*	**0.418**

underlined – native

bold – best

* - same as best (p<0.01)

dagger – better than native

- Native classifiers perform the best
- Many foreign classifiers perform as well as native
- Some foreign classifiers outperform native

Adapting from a single stream

		Accuracy		F-score		Either
Source	**Target**	**NEG**	**POS**	**NEG**	**POS**	**All**
Blogs	Blogs	4	3	3	1	7
	Reviews	1	2	2	2	4
	Twitter	1	1	3	4	7
Reviews	Blogs	2	4	5	5	10
	Reviews	5	5	5	5	10
	Twitter	2	0	3	5	9
Twitter	Blogs	5	3	5	2	8
	Reviews	3	2	2	2	5
	Twitter	5	5	4	3	10
	Best possible	5	5	5	5	10

best or same as best classifiers

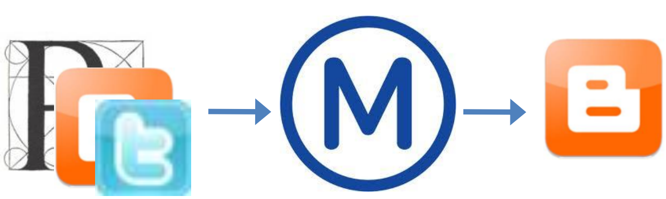
Reviews > Twitter > Blogs

Adapting from multiple streams

- Two-source mixed model
- Three-source mixed model
- Three-source voting model

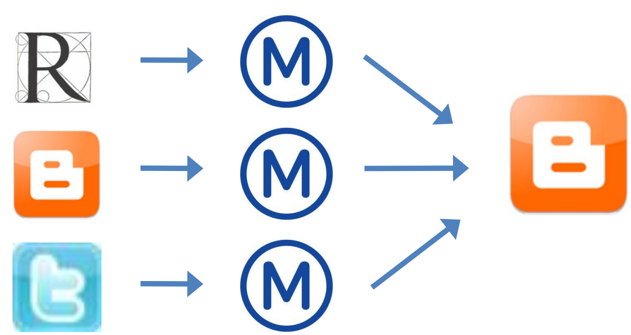
Mixed Model

train one model on documents from several sources



Voting Model

train several models, each from different source



Adapting from multiple streams

		Accuracy		F-score		Either
Source Model	Target	NEG	POS	NEG	POS	All
Mixed R + T	Blogs	3	5	5	5	10
Mixed B + T	Reviews	1	0	2	0	2
Mixed B + R	Twitter	3	2	4	4	6
Mixed all	Blogs	5	5	5	4	10
Mixed all	Reviews	5	4	5	5	10
Mixed all	Twitter	5	3	5	1	8
Voting all	Blogs	4	5	5	4	10
Voting all	Reviews	5	3	4	4	9
Voting all	Twitter	5	5	2	5	10
	Best possible	5	5	5	5	10

number of runs as good as native

Mixed == Voting

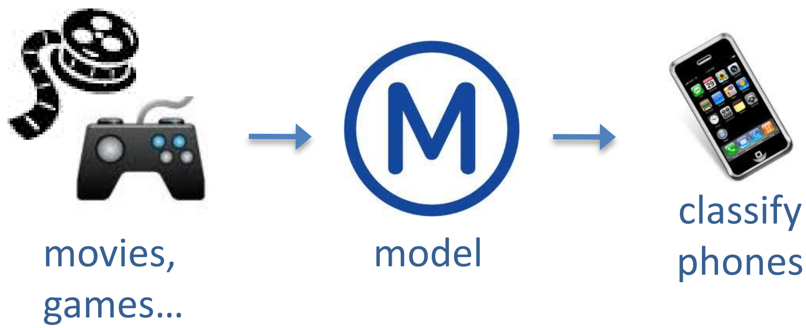
Topic-independent adaptation

Classifier	Target	Accuracy					
		Single-source			Mixed		
		Blogs	Reviews	Twitter	2 source	3 source	Voting
NEG	Blogs	**0.817**	0.732	0.773	0.788	0.801*	0.804*
	Reviews	0.662	**0.768**	0.642	0.636	0.735	0.715
	Twitter	0.791	0.762	0.862	0.816	**0.883†**	0.852*
POS	Blogs	0.628	0.683†	0.623*	0.660*	0.659*	**0.688†**
	Reviews	0.790	**0.881**	0.873*	0.821	0.881*	0.878*
	Twitter	0.620	0.448	**0.692**	0.631	0.654	0.655

Classifier	Target	Target F-score					
		Single-source			Mixed		
		Blogs	Reviews	Twitter	2 source	3 source	Voting
NEG	Blogs	0.232	0.282*	**0.325†**	0.202*	0.273*	0.256*
	Reviews	0.230	0.446	0.434*	0.304	**0.522†**	0.396
	Twitter	0.141	0.311	**0.450**	0.251	0.354	0.352
POS	Blogs	0.743	**0.835†**	0.721*	0.752*	0.747*	0.811†
	Reviews	0.880*	0.938	0.934*	0.901	**0.939***	0.937*
	Twitter	0.565*	**0.669***	0.668	0.551	0.484	0.652*

Cross-Topic Adaptation

(not in paper)



- Does not work: native << foreign
- Twitter suffers the most
  - Twitter users write differently?

Conclusions

- Streams differ in sentiment
- Cross-stream classification is possible!
- Reviews > Twitter > Blogs
- Combining streams and topics is beneficial

* within the limits of this study

More questions

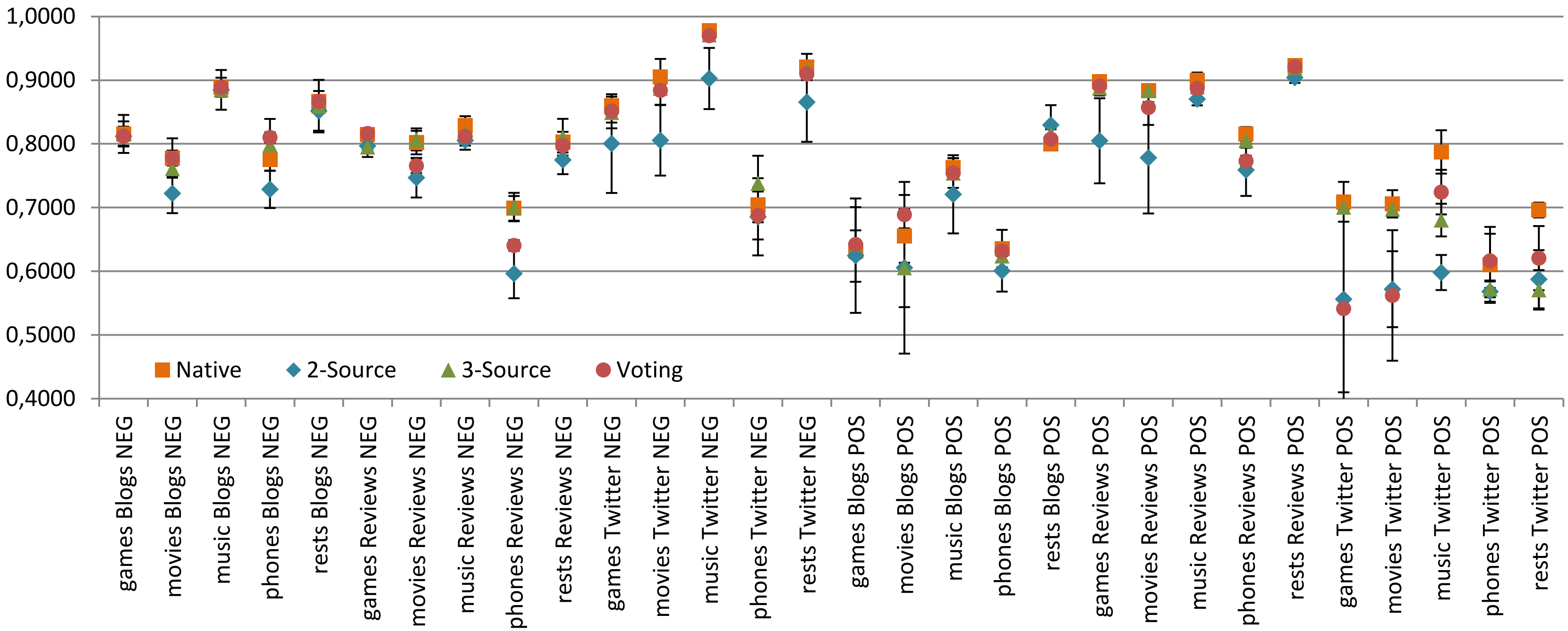
- Other topical categories
  - Politics
  - Disasters
- Multimedia websites
  - YouTube
  - Flickr
- No popularity bias

Yelena Mejova
now Yahoo! Research

Padmini Srinivasan
University of Iowa

THE UNIVERSITY OF IOWA

Adapting from multiple streams



accuracy with 99% confidence intervals