

Twitterのグラフ解析による真の有名人格付け

Using Twitter Graph Analysis to Predict Verified Accounts

イタクラ キャリー

曾根原登

Kelly Itakura

Noboru Sonehara

どんな研究？

Twitterには様々なグラフがある。よく研究されているグラフはフォロワーグラフであるが、このグラフはあまり研究されていないリツイートグラフよりあまり利便性がない。この研究では、あまり注目されていなかったリツイート、メンション、返事の三つのグラフ構造を比較研究する。それによって0.853のF-measureで、アカウントが承認済みかどうか予測することができた。

何がわかる？

フォロワー数だけが高い有名人とリツイート数などもしっかりある有名人の違いがわかる。それによってステマの有名人か本当に影響力がある真の有名人かわかる。また、有名人によって影響力の伝わり方の速さ、方向性がわかるので、状況に応じてどの真の有名人にメッセージを拡散してもらい人を動かすことができるかもしれない。

状況設定

- Twitterは、一般人だけではなく有名人も使っている。しかし一般人と有名人の境はなんだろうか？よく言われるのはフォロワー数であるが、フォロワー数は安く購入できる。リツイート数やメンション数も購入できるが、比較的高額であり、メンションされている文の内容によって偽メンションかどうか見分けが付きやすい。
- Twitterは、真似されやすいアカウントには本物かどうかをチェックして青い認証マークをつける。
- ただし一般人はたとえ真似されやすくてもアカウント認証の対象にならない。

研究内容

図1: 頻度グラフランキング上位者

Rank	Retweet Target	Mention Target	Reply Target	All Target				
1	justinbieber	7199	YouTube	18645	zaynmalik	10103	YouTube	19123
2	Tweetnesian	5035	youtube	10352	NiallOfficial	5798	justinbieber	15952
3	XSTROLOGY	3454	justinbieber	3944	justinbieber	4809	zaynmalik	13478
4	cumanNANYA	3247	Unionworld	2354	Unionworld	2714	youtube	10584
5	NiallOfficial	3086	Tweetnesian	2352	Real_Liam_Pa	2058	NiallOfficial	10006
6	UberFacts	2790	foursquare	1997	Harry_Styles	2036	Tweetnesian	7755
7	damnitstrue	2761	zaynmalik	1777	ChayLeao	1714	Unionworld	5627
8	Drakee_YMCMV	2500	K3p3Smile	1499	Louis_Tomlins	1671	cumanNANYA	4770
9	autocorrects	2220	briancohenmu	1430	null	1549	Real_Liam_Pa	4083
10	dailyteenword	1972	cumanNANYA	1388	District3music	1227	XSTROLOGY	3906
11	ihatquotes	1811	BugsBunnyZ	1328	LittleMixOffic	1166	Harry_Styles	3812
12	WhenniBoys	1798	rageofbahamu	1281	Lua_Blanco	1072	UberFacts	3527
13	AhSpeakDoang	1786	s	1262	jusxy	910	Louis_Tomlins	3479
14	SonReflexione	1665	Paul38672304	1192	sessaoextra	894	BarackObama	3087
15	BarackObama	1623	sharethis	1188	_tea	884	damnitstrue	2929
16	FILOSOFIA	1610	NiallOfficial	1122	ArianaGrande	731	ChayLeao	2861
17	zaynmalik	1598	Paul5208896	1082	JLSOfficial	678	Drakee_YMCMV	2739
18	Laughbook	1570	BarackObama	980	AustinMahone	621	AhSpeakDoang	2652
19	PEPATAHKU	1559	Etsy	930	sophiaabrahac	618	autocorrects	2404

図2: ランク対頻度グラフ

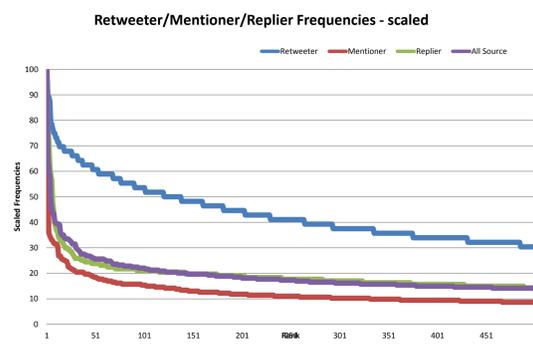
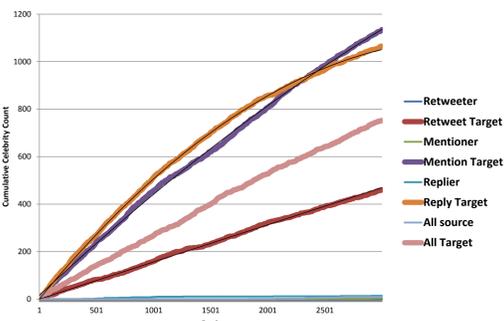


図3: ランク対有名人蓄積カウント



2012年10月17日から19日までの世界中のTweetをランダムに収集した。最終的なサイズは、2.7GBで9.7ミリオンTweetが得られた。

- そのTweetからRetweet元とRetweet先だけ抽出してRetweetグラフを作った。同様にMentionグラフとReplyグラフまたそれらをすべて総合したAllグラフを作った。
- 図1は、各グラフの中でRetweetなどされた回数が多い上位者のランキングである。黄色の者はアカウント認証されているものである。
- 図1から図4を総合するとRetweetグラフよりもMentionグラフのほうがアカウント認証者の影響力が高い。
- これらのパラメーターや、使用言語フォロワー数などを機械学習させた結果図6のように高い確率でアカウント認証者を予測できる。

図7: 相関図

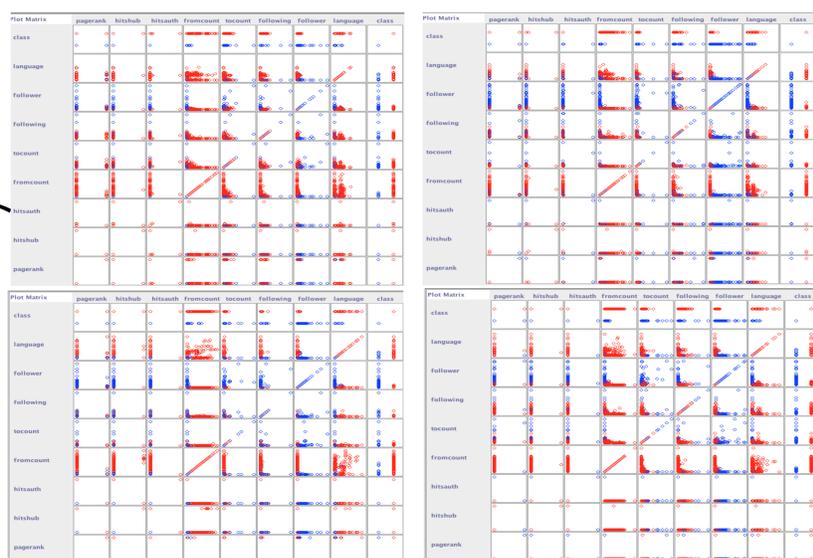


図4: 三つのグラフランキング上位者

Retweet	Mention	Reply	Merged				
zodiacohy	1.0597E-01	YouTube	3.4951E-01	null	4.0058E-04	youtube	6.4660E-04
AhSpeakDoang	7.3156E-04	youtube	2.6490E-01	James_Yammou	3.0188E-04	YouTube	6.5925E-04
autocorrects	6.1065E-04	funniest_Job	1.6623E-01	_brookez	2.4899E-04	ryeong9	4.5972E-04
ryeong9	5.5883E-04	cumanNANYA	1.1040E-01	greysonchance	2.3609E-04	AhSpeakDoang	3.8723E-04
ZonaSalahGaul	5.2428E-04	justinbieber	6.5749E-04	_2_tea	2.3609E-04	DennisErfindo	3.3282E-04
Laughbook	5.2428E-04	foursquare	6.3758E-04	ryeong9	2.3609E-04	LifeTools	3.3282E-04
LifeTools	5.2428E-04	Eminem	5.7846E-04	NiallOfficial	1.7887E-04	funniest_Job	3.2678E-04
pepatih	4.8974E-04	PerryCarting	5.1934E-04	Fazza3_RT	1.7025E-04	firstworldpics	2.9653E-04
UberFacts	4.8974E-04	sharethis	5.0020E-04	Adil_MemonM	1.7025E-04	_brookez	2.9653E-04
ihatquotes	4.8974E-04	AhSpeakDoang	4.7335E-04	home	1.7025E-04	WowSalahGaul	2.7842E-04

図5: 実験データ

	# celebrities	#commoners
Retweet Graph	117 (12.39%)	827 (87.61%)
Mention Graph	203 (33.67%)	400 (66.33%)
Reply Graph	93 (33.10%)	188 (66.90%)
Merged Graph	139 (22.75%)	472 (77.25%)

図6: 実験結果

	Precision	Recall	F-Measure
Retweet Graph	0.849	0.675	0.752
Mention Graph	0.799	0.882	0.838
Reply Graph	0.835	0.871	0.853
Merged Graph	0.849	0.77	0.808