Understanding Engagement with Insurgents through Retweet Rhetoric

Joel Nothman[1]   **Atif Ahmad**[1]   **Christoph Breidbach**[1]
David Malet[2]   **Timothy Baldwin**[1]
[1] Department of Computing and Information Systems
[2] School of Social and Political Sciences
The University of Melbourne
{jnothman,atif,christoph.breidbach,david.malet,tbaldwin}
@unimelb.edu.au

Abstract

Organisations — including insurgent movements — harness social media to engage potential consumers. They evoke sympathetic (and antipathic) response; content sharing engenders affinity and community. We report on a pilot study of presumed rhetorical intent for statuses retweeted by a set of suspected Islamic State-sympathetic Twitter accounts. This annotation is orthogonal to prior opinion mining work focused on sentiment or stance expressed in a debate, and suggests a parallel to dialogue act classification applied to retweeting. By exploring the distribution of rhetoric among Islamic State-sympathetic and general users, we also hope to identify trends in IS social media use and user roles.

1 Introduction

Social media has become an important platform for organisations and communities seeking to engage with adherents and the wider public. Through it we may follow individuals as their ideas and affiliations change, expressed through conversation, broadcast, and rebroadcast. Social scientists are keen to understand how individuals are transformed in this process of engagement: how this is effected by the organisation, and how it is realised in individual behaviour.

Recent media and scholarly studies have highlighted the use of social media by insurgent organisations. Understanding and tracking these activities is of particular interest to law enforcement and policy makers, as well as political scientists studying the nature of conflict, in terms of both monitoring and comprehending insurgent activities.

This work presents a new annotation model of partisan retweets (RTs), as "rhetorical acts". It pilots a study of content rebroadcast by suspected IS-sympathetic Twitter users. We develop an annotation schema to capture the attitude of a partisan user when retweeting content, and are able to analyse trends with respect to popularity, and transmission into/out of the IS network.

For our pilot set of suspected IS-sympathetic accounts, we find that 58% of RTs are evocative; these divide almost equally between expressing pride in the movement, expressing indignation at oppression, and transmitting religious and partisan mythology. Most others (22%) share general content, while 3% manage the ISIS Twitter network under suspension.

2 Background

Insurgent movements exploit the decentralised and colloquial nature of social media to counter mainstream narratives (Thompson, 2011; Bernatis, 2014). Berger and Morgan's (2015) seminal study of IS Twitter accounts describes their network structure and measures status sharing from within or outside the network, but gives little attention to content. Klausen (2015) analyses 10 statuses for each of 59 IS accounts, finding 40% of them deal with religious instruction, and a further 40% report from battle. Our focus on RT intent highlights the dissemination of rhetoric and its affect, classified at finer granularity than this prior work.

We investigate the construction of partisan rhetoric through distributed social media activity, while opinion mining of partisan text has largely followed Lin et al. (2006) in addressing the task of discriminating distinct points of view, in various domains (Somasundaran and Wiebe, 2009; Al Khatib et al., 2012) and granularities (Abu-Jbara et al., 2012). More recent work investigates the language used to frame discussion of contentious topics to appeal to audience values (Card et al., 2015; Baumer et al., 2015; Tsur et al., 2015), building on the premise that sentiment can be detected in statements that are not subjective or eval-

uative (Greene and Resnik, 2009). We similarly model partisan rhetorical processes that aim to engage a sympathetic audience.

Communicative units may be analysed as *dialogue acts*, which classify the intended effect on an addressee (e.g. Core and Allen, 1997). This has been applied to Twitter conversation threads with coarse classes — STATUS, REACTION, QUESTION, etc. — and with a fine-grained act hierarchy (Zarisheva and Scheffler, 2015); Zhang et al. (2011) broadly classify isolated tweets. We depart from that work to analyse rebroadcasting, not authoring, through a partisan lens.

3 Retweets as rhetoric

Propagating broadcast content has become a key feature of social media, and we choose it as a lens for analysing the IS Twitter network. Initial attempts at analysing a sample of tweets by IS-affiliated users suggest it is too noisy: the majority of statuses are poor in rhetorical and evocative content, and tend to be hard to interpret without context. In contrast, the act of propagating a status — *retweeting* in Twitter — inherently declares that it is of interest beyond its author, and usually implies that a message is encapsulated within the shared status, such that little discourse context is required to understand it. Sharing a status is a rhetorical act, although the attitude of the retweeter — our focus — often differs from that of the author.

4 Annotation schema

We examine a sample of RTs by suspected IS supporters, asking: what was the user expressing by rebroadcasting this status *assuming they are sympathetic towards IS*? We develop a shallow hierarchical schema for high coverage but reasonable robustness. At its root we distinguish between: EVOCATIVE/INSTRUCTIVE (along the lines of traditional "propaganda"); OPERATION FACILITATION; GENERAL CONTENT; and SPAM.

In some cases there is NOT ENOUGH INFORMATION to determine the category of a status. This occurs where conversational context is necessary; or where an image attached to the status was necessary, but is no longer available, frequently due to suspension of its poster.

4.1 EVOCATIVE/INSTRUCTIVE

We assume much of the content is evocative to the retweeter, as with other social media sharing (Berger and Milkman, 2012), even when it is objectively stated by the original author. We identify the following subcategories.

PRIDE: usually good news for IS, often evoking pride in IS government or land (1), military might (2)–(3), or victory (4):

(1) The building of #IS new college of medicine in ar-Raqqah #Syria *[image]*

(2) Qamishli: 4 members of the pro-Assad Maghaweer militia have defected and have now joined the Islamic State.

(3) From a small group of Jihadists surrounded in #Fallujah in 2004 into a large Islamic State that controls large parts of #Syria #Iraq in 2014

(4) BREAKING: #IslamicState shot down a warplane above kuwairs military airport !!!!! Al-Hamdulillah

INDIGNATION: expressed directly (5)–(6), or implied from news of loss (7):

(5) For all of those who normalize assad's mass killings deserve to be flayed and disemboweled alive

(6) shia rafidis slaves of kuffar, harassing sunni Muslims in Baghdad *[image]*

(7) Bismillah. iPICTURE - The U.S. organised Shi'a death squads stormed Ibn Tamiyyah mosque in Baghdad & kidnaps Sunni's *[image]*

DERISION: develops an us-vs-them dichotomy by mocking various enemies:

(8) America's track record: Successfully wiped out the Native Americans, enslaved the entire African continent, and now fighting Islam/Muslims.

(9) This is why Peshmerga cant win they need to fly all the way to Germany for treatment (This is not a joke) #IS *[image]*

INSTRUCTION: distribution of ideological materials, often religious (10), or claiming authenticity (11)–(13).

(10) The lack of showing Bara' (disavowal) towards the polytheists and apostates: Dr al-Jazouli *[url:youtube.com]*

(11) If u're a scholar in Saudi Arabia and not in jail. U're NT truthful then.

(12) How can the IS be anti-kurdish,when a large part of the attacking troops in Kobane are Kurds themselves?Don't believe the media-propaganda!

(13) #IS written by S.Qutub,taught by A.Azzam, globalized by Ben Laden, being real by Zarkawi, carried out by the 2 Baghdadi:Abu Omar & Abu Baker

4.2 OPERATIONAL

Intended to facilitate online operations, particularly maintaining Twitter network (and web sites) under adversity (14), including announcing new accounts following suspension (15):

(14) Attention bro n sisters please be careful of following people on Twitter, Trolls are changing strategy: make a pro IS account 2 lure muslims

(15) Follow ReTweet Spread n Support @AbuAdamIsBack @AbuAdamIsBack @AbuAdamIsBack

4.3 GENERAL

Creates "rapport" with followers through culture (16), humour (17), conversation (18), political news (19); or external media reports about IS and affiliates, without clear evocative aspect (20):

(16) I love how the remix isn't just thrown together. They actually put effort into making the verses go together

(17) Hahaha "pissing" myself laughing.. India launches cow urine based soda. *[url:eshowbizbuzz]*

(18) Happy Bday Sami!! @SamiAlJaber *[image]*

(19) UK spies have scanned the internet connections of entire countries: *[url:engadget.com]*

(20) Very isnightful interview with British-Pakistani Jihadi in #TTP by @Quickieleaks Must read: *[url:londonprogressivejournal.com]*

4.4 SPAM

Spam content, unrelated to IS:

(21) #workoutwednesday Back & Abs what you doing today? "Girls just wanna tank" @Bodybuildingcom *[image][url:fit-kat.com]*

(22) #GAIN_FOLLOWERS #MENTIONS #FOLLOWME & @Gamma_Monkey @jockomo141 @PATOO_S @Trans1110 @Sammi_Gemini @MREESE06 @Retweetsjp & ALL #RTS #TFB

5 Data

Affiliate accounts Prior work has painstakingly identified IS-affiliated accounts (Berger and Morgan, 2015), or has shown the success of simple heuristics (Magdy et al., 2015). The latter finds that Twitter accounts using unabbreviated forms of the IS name in Arabic-language tweets are very frequently IS supporters. This heuristic does not apply trivially to English-language tweets.

We instead combine noisy lists of suspected accounts: *LuckyTroll.club* was collected by counter-IS hacktivists on Twitter and published online,[1]

which we scraped from 2015-03-16 until 2015-05-18, yielding 36,687 accounts. Another anonymous list of 555 accounts labelled *#GoatsAgainstIsis* was published on ghostbin.com and linked from a hacktivist Twitter account. We add 36 usernames from two English-language purported IS *guide books* available from the Internet Archive (Anon., 2015a,b). Despite observing false entries — members of rival groups and unlikely Jihadis — we make no attempt to clean them.

Twitter stream Investigating IS on Twitter presents a number of challenges, particularly since Twitter began suspending affiliated accounts from mid-2014. Once suspended, Twitter's API provides no information about an account, so traditional social media analysis with follower graphs or extensive activity histories are not available. Prior work has retrieved IS user histories before their suspension, but this data is not available to us; still, we seek to make the scope of the project as broad as possible, in including both suspended and active accounts.

We use tweets collected from the Twitter Streaming API from 2014-01-01 to 2015-03-20,[2] analysed regardless of eventual suspension/retraction. An annotated status must satisfy the following criteria: (1) posted by a user in our set of suspected affiliate accounts; (2) produced using the official Twitter RT mechanism; and (3) recognised by Twitter as being in English.

We remove any duplicate RTs[3] and reduce skew to major content producers by sampling in proportion to the square root of the number of tweets by each originating author. A single annotator labelled 400 statuses with RT intent.

6 Experiments and results

Annotator agreement A second annotator, an expert in jihadist ideology, coded 100 tweets after a brief introduction to the schema. On coarse categories, the annotators agree reasonably often, $\kappa = 0.40$. This second annotator overgenerated spam labels, including various off-topic posts, e.g. news about North American weather events; conflating general and spam labels yields $\kappa = 0.45$. At the finest granularity (e.g. "religious instruction", "general humour"), agreement is a weaker $\kappa = 0.28$. Disagreement often results from content

---

	#	sus author	> 2	> 2 sus
EVOCATIVE/INSTRUCTIVE	232	56	96	16
– PRIDE	64	22	25	3
– INDIGNATION	65	10	32	6
– DERISION	15	5	6	1
– INSTRUCTION	66	12	27	4
– OTHER	22	7	6	2
OPERATIONAL	14	8	4	0
GENERAL	89	5	45	6
– ABOUT IS/AFFILIATES	7	1	2	0
– POLITICS/WARFARE	35	2	12	1
– HUMOUR	5	2	5	2
– CONVERSATION	10	0	2	0
– OTHER	32	0	24	3
SPAM	15	4	7	1
NOT ENOUGH INFO	49	17	21	10
TOTAL	400	90	173	33

Table 1: Distribution of RT intent: overall; for statuses with suspected authors; for statuses with over two sampled RTs by any/suspected users.

about groups towards which IS followers are sympathetic; one annotator saw indignation in descriptions of Gazan suffering, the second saw a general informative intent. Further schema refinement and training will hopefully reduce disagreement.

Intent distribution We analyse the RT intent distribution with respect to popularity and whether the author or retweeters are IS suspects. We regard as popular any RTs that are thrice sampled in the stream.[4] Here, suspects include those listed above, plus any accounts deactivated by 2015-09-30, often due to suspension.

Granular annotation frequencies are shown in Table 1. RTs by IS users are dominated by messages that they would find evocative or instructive (58%). Most are divided equally between pride (mostly about military strength), indignation, and instruction in group mythology. Indignation is characterised by being widely spread, beyond IS suspects, and often originating outside that network. This accords with studies showing that insurgents see themselves as addressing communal grievances (Hafez, 2007; Mironova et al., 2014). GENERAL content, often political and sourced from non-suspects, is also frequently retweeted, while a small portion (3%) of RTs maintain IS Twitter operations. Overall these distributions hint that IS RTs use religious-cultural affect and political interest as a guide towards insurgent engagement.

[4]The unknown, variable sampling rate — historically 10% of all tweets — makes this a weak heuristic.

7 Discussion

Inter-annotator agreement shows that likely intent behind an IS affiliate's RT is often determinate. Reviewing users' own remarks on their RTs might provide more robust evaluation of our annotations.[5] Suspensions make this difficult, suggesting that this task be attempted with less-controversial affiliations. We are further hampered by suspect lists collected by an unknown process that may consider the rhetoric of the user, perhaps biasing our results, e.g. derisive RTs are frequently authored and distributed by suspects.

RTs about affiliated and rival groups are among the most ambiguous for our task. Damage to a rival jihadist organisation in (23) may be a source of both indignation and pride (or schadenfreude); (24)'s apologetics for terror in the west is not clearly apologetics for IS; and though (25) literally expresses solidarity, it may pity its subject bereft of Islamic sovereignty. Such cases highlight that intent is affected by the relationship between author, retweeter and theme, suggesting future analysis akin to Verbal Response Modes (Stiles, 1992).

(23) #JabhatNusra A headquarter of #JabhatAnNusra which was bombed by the Crusaders in #Kafrdarian #Idlib (1) *[image]*

(24) #CharlieHebdo Operation wasnt a gun rampage. Gunmen had a list with targets to be assassinated rest of the staff & civilians were free to go

(25) Oh Allah bless our brothers in the UK who hav held the rope of haq.. Even in difficult tyms and never compromised.. Ya Allah bless them...

8 Conclusion

We have presented an initial treatment of the act of social media rebroadcasting as akin to a speech act, laden with rhetorical intent in the context of partisan propaganda. We hope our work lights the way towards a more general model of this quintessential social media communicative act. Though we leave automatic classification to future work, large scale analysis of IS RT intent may allow us to analyse different types of IS-affiliated users, and identify changes in rhetoric over time and place that are indicative of radicalisation.

[5]Since mid-2014, Twitter allows users add remarks when retweeting. The original status is not provided on the stream in such cases, and so is inaccessible after suspension.

References

Amjad Abu-Jbara, Pradeep Dasigi, Mona Diab, and Dragomir Radev. 2012. Subgroup detection in ideological discussions. In *Proceedings of the 50th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 399–409.

Khalid Al Khatib, Hinrich Schütze, and Cathleen Kantner. 2012. Automatic detection of point of view differences in Wikipedia. In *Proceedings of COLING 2012*, pages 33–50.

Anon. 2015a. *Hijrah to the Islamic State*. The Internet Archive. https://archive.org/details/GuideBookHijrah2015-ToTheIslamicState.

Anon. 2015b. *The Islamic State*. The Internet Archive. https://archive.org/details/TheIslamicState2015-FullEbook.

Eric Baumer, Elisha Elovic, Ying Qin, Francesca Polletta, and Geri Gay. 2015. Testing and comparing computational approaches for identifying the language of framing in political news. In *Proceedings of the 2015 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 1472–1482.

J.M. Berger and Jonathon Morgan. 2015. *The ISIS Twitter census: defining and describing the population of ISIS supporters on Twitter*. Number 20 in The Brookings Project on U.S. Relations with the Islamic World. Brookings Institution.

Jonah Berger and Katherine L. Milkman. 2012. What makes online content viral? *Journal of Marketing Research*, 49(2):192–205.

Vincent Bernatis. 2014. The Taliban and Twitter: Tactical reporting and strategic messaging. *Perspectives on Terrorism*, 8(6).

Dallas Card, Amber E. Boydstun, Justin H. Gross, Philip Resnik, and Noah A. Smith. 2015. The media frames corpus: Annotations of frames across issues. In *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (Volume 2: Short Papers)*, pages 438–444.

Mark Core and James Allen. 1997. Coding dialogs with the DAMSL annotation scheme. In *Working Notes of the AAAI Fall Symposium on Communicative Action in Humans and Machines*, pages 28–35.

Stephan Greene and Philip Resnik. 2009. More than words: Syntactic packaging and implicit sentiment. In *Proceedings of Human Language Technologies: The 2009 Annual Conference of the North American Chapter of the Association for Computational Linguistics*, pages 503–511.

Mohammed Hafez. 2007. *Suicide Bombers in Iraq: The Strategy and Ideology of Martyrdom*. United States Institute of Peace Press.

Jytte Klausen. 2015. Tweeting the Jihad: Social media networks of western foreign fighters in syria and iraq. *Studies in Conflict & Terrorism*, 38(1):1–22.

Wei-Hao Lin, Theresa Wilson, Janyce Wiebe, and Alexander Hauptmann. 2006. Which side are you on? Identifying perspectives at the document and sentence levels. In *Proceedings of the Tenth Conference on Computational Natural Language Learning (CoNLL-X)*, pages 109–116.

Walid Magdy, Kareem Darwish, and Ingmar Weber. 2015. #FailedRevolutions: Using Twitter to study the antecedents of ISIS support. *arXiv CoRR*, abs/1503.02401.

Vera Mironova, Loubna Mrie, and Sam Whitt. 2014. The motivations of Syrian Islamist fighters. *CTC Sentinel*, 7(10):15–17.

Swapna Somasundaran and Janyce Wiebe. 2009. Recognizing stances in online debates. In *Proceedings of the Joint Conference of the 47th Annual Meeting of the ACL and the 4th International Joint Conference on Natural Language Processing of the AFNLP*, pages 226–234.

William B Stiles. 1992. *Describing talk: a taxonomy of verbal response modes*. SAGE Series in Interpersonal Communication. SAGE Publications.

Robin L. Thompson. 2011. Radicalization and the use of social media. *Journal of Strategic Security*, 4(4):167–190.

Oren Tsur, Dan Calacci, and David Lazer. 2015. A frame of mind: Using statistical models for detection of framing and agenda setting campaigns. In *Proceedings of the 53rd Annual Meeting of the Association for Computational*

Linguistics and the 7th International Joint Conference on Natural Language Processing (Volume 1: Long Papers), pages 1629–1638.

Elina Zarisheva and Tatjana Scheffler. 2015. Dialog act annotation for Twitter conversations. In *Proceedings of the 16th Annual Meeting of the Special Interest Group on Discourse and Dialogue*, pages 114–123.

Renxian Zhang, Dehong Gao, and Wenjie Li. 2011. What are tweeters doing: Recognizing speech acts in Twitter. In *Analyzing Microtext: Papers from the 2011 AAAI Workshop*.