Understanding Russian information operations using unsupervised multilingual topic modeling

Peter A. Chew and Jessica G. Turnley

Galisteo Consulting Group, Inc., 4004 Carlisle Blvd NE Suite H, Albuquerque, NM 87107
{pachew,jgturnley}@galisteoconsulting.com

Abstract. What does this or that population think about a given issue? Which topics 'go viral' and why? How does disinformation spread? How do populations view issues in light of national 'master narratives'? These are all questions which automated approaches to analyzing social media promise to help answer.

We have adapted a technique for multilingual topic modeling to look at *differences* between what is discussed in Russian versus English. This kills several birds with one stone. We turn the data's multilinguality from an impediment into a leverageable advantage. But most importantly, we play to unsupervised machine learning's strengths: its ability to detect large-scale trends, anomalies, similarities and differences, in a highly general way.

Applying this approach to different Twitter datasets, we were able to draw out several interesting and non-obvious insights about Russian cyberspace and how it differs from its English counterpart. We show how these insights reveal aspects of how master narratives are instantiated, and how sentiment plays out on a large scale, in Russian discourse relating to NATO.

Keywords: information operations, topic modeling, multilingual, Russia

1 Introduction

In the current geopolitical climate, and with growing use of social media, there is widespread interest in questions such as: What does this or that population think about a given issue? Which topics go viral and why? How does disinformation spread? How do populations view issues in light of national or religious 'master narratives'? The vast open-source trove that is social media has not only pushed these questions high on the agenda; it also begs automated solutions to providing answers.

Most approaches, such as sentiment analysis (2), (3), (7), deception detection (6), topic modeling (4), or master narratives analysis (8), focus in narrowly on one or another of these questions. In this paper, we describe an approach that attempts to balance many of these issues simultaneously. That may sound like a grand claim, but to be clear, we are not trying to subsume all the achievements of previous, focused work. Instead, we outline a role we think is eminently suited to *unsupervised machine learning* techniques: allowing top-down exploration of multilingual social media not

adfa, p. 1, 2011.

just to find trending topics cross-linguistically, but also to aid understanding of *differences* between different important subsets of the data, for example the subsets of social media data represented by English- versus Russian-language posts. Via examples, we show how exploring such differences can lead to significant, non-obvious insights about how different populations view the world. Compared to most existing work, our unsupervised approach essentially reallocates the labor between human and machine: the machine does what it is best at – finding patterns, similarities and differences, leaving the higher-order analysis to the human, but making the human's task easier by focusing the human's attention in first and foremost on the most material patterns and prominent in the data, and helping a human avoid getting 'lost in the weeds'.

This paper is structured as follows. In section 2 we briefly describe STEMMER, our approach to multilingual topic modeling, explaining how it can be adapted to review differences between subsets of the data ('information spaces'). Section 3 presents the results of applying this approach to two Twitter datasets, each comprising over 100,000 English-, Russian- and Ukrainian-language posts. We show a number of examples of non-obvious, and potentially significant and actionable, insights that our approach allows us quickly to tease out of the data. Finally, we conclude in section 4.

2 STEMMER: a framework for unsupervised analytics

We examine unsupervised pattern recognition on multilingual social media, and do not pursue supervised or similar analyses in this paper. We call the unsupervised-analysis approach that we have tailored specifically to multilingual data STEMMER (System for Top-down Exploration of Mixed Multilingual Electronic Resources), described in detail in (5). STEMMER, a modified version of standard Latent Semantic Analysis (LSA), differs from other topic modeling approaches in that it induces *cross-language* topics, as in Figure 1 (ibid). LSA takes any collection of text and derives prominent topical patterns from that text deterministically: the same input always produces the same output, and the top *n* topics are guaranteed to be the *n* most prominent topics in the corpus. This property qualifies LSA eminently well as a 'top-down' approach, one that helps analysts not miss the forest for the trees. STEMMER adds a linear-algebra pre-processing step, optimized for LSA, shoe-horning all documents into a single multilingual space. Recommending STEMMER is that it has been empirically validated: by fine-tuning, we have brought its accuracy up to 94.8% (9).

Doc #	**Source text**
10433	... RT @AFP: UPDATE: 2,000 Russian soldiers land in \
570	Russia seizes control of Crimea! .... #Ukraine http://t.co/hDjGFvOfoc
8082	Casi 2,000 soldados rusos han aterrizado en las ultimas horas en Crimea…

Fig. 1. A topical pattern from multilingual Twitter data

In the work cited (11, 15) the focus is on using STEMMER to identify groups of topically similar documents, where similarity is determined ultimately by the words used in each document. This is in itself a useful function for analysts, because it helps

analysts sort the data into 'buckets' that make sense, and again not miss the forest for the trees. But STEMMER can also be used to look at *differences* between different areas of the 'forest'. Most digital data, including Twitter, has an abundance of pre-supplied or easily inferable metadata, e.g. geospatial coordinates, language, author, date and time, etc. This metadata supplies obvious ways to subdivide the data.

Guided by how the Russian government talks about 'information spaces', we choose here to look at gross differences by *language*. Framed this way, the problem becomes one which unsupervised learning, and STEMMER in particular, is ideally placed to handle. We can use STEMMER to detect the biggest topical *differences* between Russian and English. Technically, the approach to finding such high-level topical differences is straightforward, once we have the STEMMER framework in place. A key output of STEMMER is a document (Twitter post) by topic matrix. Each cell in that matrix encodes how 'strong' a given topic is in a given document. Since we know the metadata (its language) for each document a priori, we can subdivide the matrix into blocks by language, as shown in Fig. 2. Then, it is simple, intuitive, and justifiable in linear-algebraic terms, to calculate, for each topic, a 'strength' of that topic by language: to do this, we can calculate for each block the sum of the squared cell values (in linear algebra terms, the magnitude) in that block.

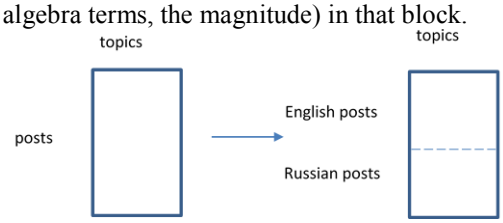

Fig. 2. Subdivision of document-by-topic matrix into blocks

The final step is to sort the topics by 'magnitude variance', as shown in Table 1. We can then explore the topic further by looking at what terms and documents most characterize that topic. This gives insight into *how*, topically, differences between one information space and another manifest themselves.

Topic #	English Magnitude	Russian Magnitude	Magnitude variance
1	0.0653	0.1372	0.0719
2	0.0825	0.1261	0.0436
…	...	...	...

Table 1. Topics sorted by 'language magnitude variance'

3 Results

We used the approach above to explore two separate datasets, each a sample of posts from the Twitter 'firehose' and within specified date ranges, shown in Table 2.

We used STEMMER to analyze the top 90 cross-linguistic topics for each dataset. We use a multilingual parallel corpus for the pre-processing step to 'multilingualize' each post, a process described in detail in (9). We manually augmented that parallel

corpus with the top 100 most frequent out-of-vocabulary words in each language, a process that takes an hour or two for someone with relevant language expertise. The ability of STEMMER to return multilingual topics is not absolutely dependent on this kind of manual intervention but is nonetheless improved with a relatively small investment of effort. 'Out-of-vocabulary words' tend to be proper names (e.g. Putin / Путин) or technical terms (e.g. генсек, Russian acronym for 'general secretary').

#	Dataset description	Date range	Seed words	# of posts by language		TOTAL
1	'Brilliant Jump' NATO exercises	May-June 2016	NATO, HATO	EN RU	127,220 79,514	**206,734**
2	NATO discussion in 2016 US election	10/13/2016-11/28/2016	NATO, HATO	EN RU	81,969 79,465	**161,434**

Table 2. Datasets analyzed

3.1 'Brilliant Jump' NATO exercises dataset

Language weighting plausibility: Topic #6, highly weighted towards English, included as top keywords: Корею (Korea), blast, Korea, Клинтон (Clinton), north, Clinton, речь (speech), speech. Twitter posts highly representative of Topic 6 were:

- Clinton To Blast Trump On North Korea, NATO In Foreign Policy Speech 😂😂😂
- Reuters: Clinton to blast Trump on North Korea, NATO in foreign policy speech

From just the above it should be clear the common thread in this topic is Clinton and Trump's discussion of North Korea with respect to NATO. It is plausible that in the context of a dataset from May-June 2016, this topic is 'weighted toward' the English-language information space, for no other reason than that relatively few Russian speakers had much interest in the US Presidential election this early in the campaign.

Similarly, topic #5 was one of those most highly weighted towards Russian, we believe for a similar reason. Top keywords for this topic were: secretary, Poroshenko, Порошенко (Poroshenko), советником (advisor), генсек (secretary), ex, экс (ex), генсека (secretary), назначил (appointed). An English-language Twitter post most representative of this topic was 'Poroshenko appointed the ex-secretary of NATO as his advisor'. Poroshenko, the President of Ukraine, is relatively unknown to English speakers and it is thus plausible that few English-speaking Twitter users discuss him.

Result of interest: Topic #11 (Russian weighting .1152, English weighting .0992) was also among the topics with greatest magnitude variance, and included these top keywords: Russia, Poland, strike, global, глобального (global), удара (strike). A Twitter post highly weighted in Topic 11 was '#News. In the [Russian] Federation Council an announcement was made about a "global strike" by NATO on Russia /#Russia.'* (asterisks here and below denote our translation from original Russian).

With some background knowledge of NATO activities, we were aware that 'global strike' is a defensive/deterrent capability that NATO developed to strike back any-

where in the world within an hour against an emerging threat. Whether as a result of deliberate disinformation, a misconstruing of the true nature of 'global strike', attempts by Kremlin-controlled trolls to whip up fear, or something else, the Russian-language posts that are most representative of topic 11 seem to misinterpret 'global strike' as a direct threat to Russia, an intent of NATO to attack Russia. Further, with some background knowledge of Russian discourse and use of master narratives, we can say that this may be an instantiation of the 'Fortress Russia' master narrative that claims that Russia is under global threat by outsider adversaries (10).

To verify that this result was a real phenomenon in the data, we counted the posts in the dataset mentioning both 'global' and 'strike' (and inflected forms in Russian), by language. It turned out that only 78 English posts mentioned the two words, while 2,541 Russian posts did, confirming the reality of the phenomenon STEMMER automatically found as an emergent property of the data.

3.2 2016 Presidential election dataset – Results of interest

Two topics that we thought were of most interest in this dataset were #25 (weighted towards Russian) and #22 (weighted towards English). Top keywords for topic #25 include 'Syria' and 'Lavrov'. A representative Twitter post highly weighted in this topic was 'Lavrov called NATO troops' bombing of Yugoslavia aggression'*.

It was initially unclear to us what the 1990s NATO bombing of Yugoslavia had to do with Syria (a top keyword for this topic) and 2016. However, a quick internet search led us to https://www.youtube.com/watch?v=kSXaAU-szqw, a short interview where Sergei Lavrov (Russian foreign minister) argues Russia's bombing of Aleppo, Syria, is no different from NATO's bombing of Yugoslavia. So here, STEMMER helped direct us to part of the Russian narrative of which we were unaware.

Finally, topic #22, weighted towards English, contained numerous posts similar to the following: 'While you were focused on the Walking Dead, NATO and U.S. marines prep for Russian war'. We were again unsure of the connection between NATO, Russia and the 'Walking Dead'. An internet search for 'walking dead NATO' turned up sites like thefreethoughtproject.com, russia.trendolizer.com, and thefringe-news.com. When we opened some of these, unexpected pop-ups led us to suspect strongly the sites might be infected with malware and so we quickly curtailed further investigation. Were these sites themselves part of the Russian multi-pronged information warfare strategy? Possibly – and this too could be of interest to an analyst.

4 Conclusion

In this paper we outline a way to turn the multilinguality of social media content, usually an obstacle for analysis, into an asset; and at the same time a way also to take Russian strategic thinking as a starting-point for deconstructing Russian 'information warfare' so that analysts can better understand its vectors of attack and thus be in a better position to develop countermeasures. The method is top-down, focusing first on the most important trends and patterns, and also on the most important *differences*

between different natural subsets of the social media universe, for example Russian versus English content. The value of our approach is that it helps an analyst quickly see what is most important in hundreds of thousands of posts without being burdened by reading and translating many individual posts – essentially, it helps analysts see the forest for the trees. This technique has been empirically validated, although our focus in this paper has been not on the empirical validation which is discussed elsewhere, but on demonstrating the usefulness and plausibility of its results by example.

We believe the approach we have taken here points the way to best practices in data analytics of this type: keep the human fully engaged in the loop, and cleanly separate between areas of responsibility for human and computer: have the machine do what it is best at – finding patterns, and groups of things that are 'similar' or 'different' – but let the human effectively focus his energies on what humans do best: relating those patterns to subject-matter expertise, such as prior knowledge of country-specific dynamics. This, we believe, is the best way to ensure that neither the human nor the computer are allowed to be sidetracked by confirmation bias or other kinds of systemic bias. The human's attention should always be kept focused on what is most important *within a particular dataset*, and our approach is designed to do just that.

5 References

1. Duda, Richard O., Hart, Peter E., and Stork, David G. 2001. Unsupervised Learning and Clustering. *Pattern classification* (2nd ed.). Wiley. ISBN 0-471-05669-3.
2. Kim, S.-M., and Hovy, E. 2004. Determining the sentiment of opinions. In *COLING '04 (Proc. of the 20th Intl Conference on Computational Linguistics)*, pp. 1367–1373.
3. Pang, B., Lee, L., and Vaithyanathan, S. 2002. Thumbs up? Sentiment Classification using Machine Learning Techniques. *Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP)*, Philadelphia, July 2002, pp. 79-86.
4. Bader, Brett W., Berry, Michael W., and Browne, Murray. 2008. Discussion Tracking in Enron Email Using PARAFAC. In *Survey of Text Mining* II, eds. Michael W. Berry and Malu Castellanos. Springer: London. pp. 147-163.
5. Chew, Peter A. 2015. 'Linguistics-Lite' Topic Extraction from Multilingual Social Media Data. *Social Computing, Behavioral-Cultural Modeling, and Prediction* (Vol. 9021, LNCS), eds. Nitin Agarwal, Kevin Xu, Nathaniel Osgood. pp 276-282.
6. Tsikerdekis, Michail and Zeadally, Sherali. 2014. Online deception in social media. In *Communications of the ACM* Volume 57 Issue 9, pp. 72-80.
7. Center for Computational Analysis of Social and Organizational Systems. 2016. Multilingual Twitter Sentiment Analysis. Accessed 07/27/2016 at http://www.casos.cs.cmu.edu/projects/projects/mltsa.php.
8. Halverson, J., Corman, S., and Goodall, H. 2011. Master Narratives of Islamist Extremism. New York: Macmillan.
9. Chew, Peter. 2016. Multilingual Retrieval and Topic Modeling using Vector-Space Word Alignment. Galisteo Consulting Group, Inc. Technical Report GCG002, February 2016. DOI: 10.13140/RG.2.2.21482.11205.
10. Bouveng, Kerstin. 2010. The Role of Messianism in Contemporary Russian Identity and Statecraft. Durham Theses, Durham University. Accessed at http://etheses.dur.ac.uk/438.