

ISIS in the Eyes of the Dutch

Bas Hendrikse¹, Mena B. Habib², and Maurice van Keulen³

¹ University of Twente, The Netherlands

`b.hendrikse@student.utwente.nl`,

² Maastricht University, The Netherlands

`m.habib@maastrichtuniversity.nl`,

³ University of Twente, The Netherlands

`m.vankeulen@utwente.nl`

Abstract. The presence of militant group Islamic State of Iraq and Syria (ISIS) is growing. Terrorist attacks in Europe and an incoming stream of refugees in the south of the continent are some of the reasons Europe is getting socially involved in the Middle-Eastern war. It might seem that the Netherlands could become a target of the organization too. But are Dutch citizens concerned about this? In this paper, we describe the reaction of the Dutch on ISIS by analyzing what they say on Twitter about the organization. With the use of text classification, topic modeling and visualization tools, we were able to retrieve Tweets about ISIS and create a network graph displaying the ten main topics about ISIS which Dutch people tweeted about, in addition to a word cloud, displaying the words which were most used in the Tweets. The visualizations are used to analyze what topics people discussed most regarding ISIS on Twitter and how they felt about these topics. Understanding the social network responses to ISIS' attacks can give us some clues to understand its impact on the society.

1 Introduction

1.1 Problem Statement

Many countries are involved in the chaos which is happening in the Middle-East. They are not only fighting ISIS in Syria and Iraq, but are fighting ISIS in their own countries too. Refugees who seek shelter are mixed with the local inhabitants, which lead to quite some disturbance in Europe. In combination with terrorist attacks in Paris in November 2015⁴ and Brussels in March 2016⁵, and with the use of propaganda videos [4], people may fear being the next target of the terrorist organization.

1.2 Motivation

The main goal of this research is to find out what Dutch people think about the militant group ISIS. With the use of the opinions posted on Twitter, we measure

⁴ Live blog about the Paris attack: <https://goo.gl/XoqzSX>

⁵ News article about the Brussels attack: <https://goo.gl/76tMXU>

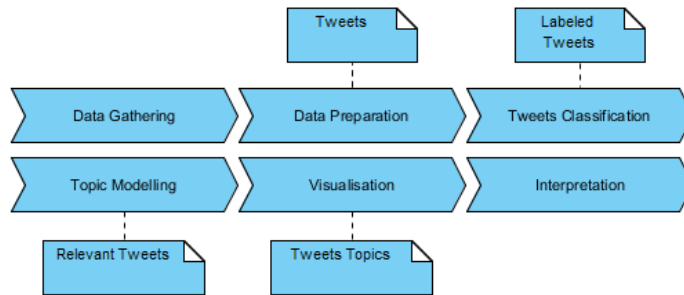


Fig. 1: Our proposed approach

the effectiveness of the actions ISIS undertakes to spread terror in Europe. We learn which words are used most often in combination with keywords identifying the militant group ISIS. With the use of these words, we can create an image of the impact of ISIS' actions on the sentiment of the opinions. The result of this research could be used to identify steps which can be undertaken to make Dutch citizens feel safer and give an overview on the impact of terrorist attacks. Although Twitter is not representative of the whole society, it still can serve as an indicator for what people think about some topic. This results into the following research question: *What do Dutch people on Twitter say about ISIS?*. To answer this question, two sub questions are defined. The first one is: What are the main topics people discussed regarding ISIS? And what did people say about the topics? The second one is: What words do people use to describe their feelings about ISIS?

2 Approach

To find out what people say and how people feel about ISIS, we processed and analysed 318,049 Dutch Tweets from the years 2014 and 2015. The Tweets needed to be processed by several steps before they could be analyzed. An overview of the processes we developed can be found in Figure 1.

2.1 Data Gathering

A dataset was set up with Twitter messages from the years 2014 and 2015. This dataset is continuously being collected by E.T.K. Sang and A. van den Bosch [7]. The Twitter messages (Tweets) were filtered on the following keywords identifying ISIS: 'ISIS', 'ISIL', 'Daesh', 'Daish', 'Islamic State' and 'Islamitische Staat' ('Islamic State' in Dutch). Our resulting dataset contained 318049 Tweets.

We noticed that a lot of Tweets were *Retweets*, a reaction of another user on a Tweet or just a duplication of the Tweet by another user, which contained most of the Tweet they were responding to. These duplicates would pollute the data which would be used for the analysis of the Tweets. For this reason, we filtered out all Retweets from our dataset.

We also noticed that many Tweets are representing news headlines posted by news providers. As our goal focuses on what people say and think about ISIS, we decided to filter out Tweets showing news headlines. For this reason, we removed all Tweets containing a URL from the dataset, as most Tweets with news headlines contained a link to an article.

2.2 Tweets' Classification

We used a Naive Bayes classifier [5] provided by the Machine Learning for Language Toolkit (MALLET) [6] to filter out Tweets that do not refer to the terrorist organization ISIS. All Tweets are evaluated by the trained classifier, the Tweets which are not discussing the terrorist organization were not included in the new dataset.

To test the classifier we use two different sets of data, test set 1 consists of 1847 Tweets automatically filtered on "Daesh" and labeled as True. The accuracy of this test set was 0.98, the recall for Tweets labeled as True was 0.98, the precision was 1.0 and the F1 was 0.99. Test set 2 consists out of 14 Tweets manually filtered on "Isis" and labeled as True and 30 Tweets manually filtered on "Isis" and labeled as False. The accuracy of this test set was 0.82, the recall for Tweets labeled as True was 0.714, the precision was 0.714 and the F1 was 0.714. For the same test set for Tweets labeled as False, the recall was 0.86, the precision was 0.86 and the F1 was 0.86.

2.3 Topic Modeling

The irrelevant Tweets which were not about the terrorist organization have been filtered out in the Tweet classification step. The next step is to interpret all the Tweets. To do so, we used topic modeling approach. The topic modeling process groups the Tweets which have a common subject.

MALLET Gibbs sampling [3] was used to perform the Topic Modeling.

We used all the data collected and processed so far as an input for the topic modeling process. A total of 10 topics were created by MALLET, each topic is described by 20 keywords which are most common for the topic. We analyzed the keywords which were created for each topic and interpreted the subject for each topic. We interpreted the first topic for example as a topic with the subject 'Dutch Politics', we chose this topic because we recognized names of political parties in the keywords. Other subjects included 'ISIS in the media', 'Attacks in Iraq and Syria', 'ISIS and the Islam', 'Compared to other terrorist organisations', 'Naming of ISIS', 'Refugees', ' Hamas in Israel', 'Demonstration against ISIS in The Hague' and 'Support for and against ISIS from countries'.

It is not possible to directly see what people think about ISIS by only looking at the subject of the topic. This is why we linked a sample of 1000 Tweets from the classification process to the topics, which might give an impression what they think about the topic and therefore how they feel about ISIS' actions.

2.4 Visualization

With the use of the previously described methods and with the use of the data visualisation tool Gephi [2], a network graph is created which shows the relationship between a sample of Tweets and the most common topics. We created two types of nodes; the first type of nodes were the topics, the second type of nodes were the Tweets themselves from the sample. The Tweets are connected to the principal topic of the Tweet. The size of the topics is determined by the number of connected Tweets. The resulting network graph can be found online.⁶

Another way to get to know what people say about ISIS is to analyze what terms they use most often in their Tweets about ISIS. By creating a word cloud, we present a visual representation of what words are most used in general in all Tweets. With the use of this data, we could learn what people say about ISIS and what the concerns of Dutch citizens are about ISIS and hence answer the second research question.

The word cloud was created with the use of the web application Voyant Tools⁷, a web-based reading and analysis environment for digital texts [8]. As a preprocessing, we tokenized the Tweets and removed the Dutch stopwords. In addition to the stopwords, we also filtered out more terms which are too dominant in the dataset but are not much saying.

The resulting word cloud can be found in online.⁸

3 Results Discussion

3.1 The network graph

The network graph displays ten topics which are the most talked about on Twitter in Dutch as the primary nodes. The nodes are connected to a sample of 1000 Tweets about ISIS. The topics were formed in the topic modeling process. We would like to confirm whether the topics have been correctly labeled and as a next step want to list the topics that are most controversial on Twitter. With the use of this validation we can answer the first sub research question.

We selected five topics we expect to be most related to ISIS threat in The Netherlands. Other topics (for example the topic about the name of ISIS) are less relevant, it is less likely people will Tweet their feelings about ISIS when referring to this topic.

Most of the Tweets about the Dutch Politics in the graph are opinions on how the government should act against ISIS. Many Twitter users who Tweet about this topic do not agree with the point of view of some political parties.

⁶ The network graph can be found here: http://bashendrikse.nl/files/SoMePeAS/Network_Graph.png

⁷ <https://voyant-tools.org/>

⁸ The word cloud can be found here: http://bashendrikse.nl/files/SoMePeAS/Word_Cloud.png

Tweets linked to ISIS in the media were about topics covered in TV shows and on the radio. The topic also covers Tweets that are about the propaganda ISIS is spreading via the Internet.

It is notable that that largest part of Tweets linked to the topic named 'Refugees' are not about refugees. People give their diverse opinions on ISIS with this topic. The topic is ambiguous as the Tweets which are linked to the topic are about diverse subjects.

The topic about Hamas in Israel is scrambled, there is not one common subject. For example, people Tweet about what political parties say about ISIS, they also Tweet about their anger for ISIS but also for the disappointment that ISIS misuses religion for their deeds. There are many other things people Tweet about.

Almost all of the Tweets which are linked to the topic labeled as 'Demonstration *against* ISIS in The Hague' are indeed about the demonstration in The Hague. But people talk about a demonstration in favor of ISIS, not against ISIS. The topics should therefore be labeled as Demonstration *for* ISIS in The Hague.

3.2 Word cloud analysis

The word cloud shows the words which were used most often in the Tweets about ISIS. The size of each word corresponds to the number of occurrences in the Tweets, the largest words were mentioned the most and the smallest words were mentioned the least.

The two largest words, 'islam' and 'moslims', are about religion, many Tweets in the network graph are discussing the idea that ISIS is doing their actions in the name of the Islam. There is a strong link between the topics Islam and ISIS in the Tweets. In addition to 'The Netherlands', there are mentions for other countries too, like 'Iraq' and 'Syria'. The Tweets are probably discussing the war with ISIS in those countries there. But 'Turkey' is also mentioned. Turkey is the border country of Syria and Iraq, so probably has to fight against ISIS too. It also has to deal with the incoming refugees [1]. Another big word is the word 'Koerden' (Kurds in English), which is a large ethnic group in the Middle East. Although the Kurds are not one unified group, the Kurds were fighting against ISIS [9], which could be the reason people tweeted about it. There are also entities involving rulers of the countries, words like 'Assad', the president of Syria and 'Erdogan', the Turkish president. But there are also terms discussing Dutch parties, like 'pvv', 'geertwilderspvv', 'pvda' and 'vvd'.

The rest of the terms give a good impression what the situation is with ISIS. A number of words describe the war with ISIS and violence which comes with it. The names of countries and parts of the world could be about the involvement of other countries in the war, where words are about politicians which say something about ISIS and the war. Words like 'demonstratie', 'vlag', 'schilderswijk', 'haag' could be about the demonstration in The Hague, just like was mentioned in the network graph analysis.

4 Conclusions

In this paper, we focused on finding what people on Twitter say about ISIS. By using some Text Mining approaches, we found out what the main topics were people discussing regarding ISIS on Twitter. With the use of Text classification, Topic Modeling and visualization tools, we were able to filter out Tweets about ISIS and were able to create a network graph displaying the ten topics about ISIS on which Dutch people tweeted about and a word cloud displaying the words which were most used in the Tweets. The visualizations were used to analyze what topics people talked most about regarding ISIS on Twitter and how they felt about the topics.

The main topics people tweeted about were: Dutch politics, ISIS in the media, attacks in Iraq and Syria, ISIS and the Islam, other terrorist organizations, the name of ISIS, refugees, a demonstration for ISIS in the Hague and support for and against ISIS from countries. By looking at the Tweets which were connected to topics which were closest related to ISIS in The Netherlands, we found that people do not tell that they are scared or angry, but are often more disappointed in the actions of ISIS. Most of the Tweets are not about their expression of a certain emotion, but often just make a statement on what happened.

In our future work, we plan to extend our study to include Tweets from 2016 where more ISIS crimes took place in Europe. We plan also to do a deeper sentiment analysis to check how these crimes affect the Dutch mode.

References

1. F. Baban, S. Ilcan, and K. Rygiel. Syrian refugees in turkey: pathways to precarity, differential inclusion, and negotiated citizenship rights. *Journal of Ethnic and Migration Studies*, pages 1–17, 2016.
2. M. Bastian, S. Heymann, and M. Jacomy. Gephi: An open source software for exploring and manipulating networks. 2009.
3. T. Griffiths. Gibbs sampling in the generative model of latent dirichlet allocation. 2002.
4. A. Hatton and M. Nielsen. "war on terror" in our backyard: effects of framing and violent isis propaganda on anti-muslim prejudice. *Behavioral Sciences of Terrorism and Political Aggression*, pages 1–14, 2016.
5. C. D. Manning, P. Raghavan, and H. Schütze. *Introduction to Information Retrieval*. Cambridge University Press, New York, NY, USA, 2008.
6. A. K. McCallum. Mallet: A machine learning for language toolkit. <http://mallet.cs.umass.edu>, 2002.
7. E. T. K. Sang and A. van den Bosch. Dealing with Big Data: the Case of Twitter. *Computational Linguistics in the Netherlands Journal*, 3:121–134, 2013. ISSN: 2211-4009.
8. S. Sinclair, G. Rockwell, and the Voyant Tools Team. Voyant tools (web application). <http://voyant-tools.org/>, 2012.
9. B. Smith and R. Page. *ISIS and the sectarian conflict in the middle east*. 2015.