Proceedings of the Thirteenth International AAAI Conference on Web and Social Media (ICWSM 2019)

FA-KES: A Fake News Dataset around the Syrian War

Fatima K. Abu Salem,[1]* Roaa Al Feel,[1] Shady Elbassuoni,[1] Mohamad Jaber,[1] May Farah[2]
[1]Computer Science Department
American University of Beirut
[2]Sociology and Anthropology Department
American University of Beirut
fa21@aub.edu.lb, rba15@mail.aub.edu, se58@aub.edu.lb, mj54@aub.edu.lb, mf15@aub.edu.lb

Abstract

Most currently available fake news datasets revolve around US politics, entrainment news or satire. They are typically scraped from fact-checking websites, where the articles are labeled by human experts. In this paper, we present FA-KES, a fake news dataset around the Syrian war. Given the specific nature of news reporting on incidents of wars and the lack of available sources from which manually-labeled news articles can be scraped, we believe a fake news dataset specifically constructed for this domain is crucial. To ensure a balanced dataset that covers the many facets of the Syrian war, our dataset consists of news articles from several media outlets representing mobilisation press, loyalist press, and diverse print media. To avoid the difficult and often-subjective task of manually labeling news articles as true or fake, we employ a semi-supervised fact-checking approach to label the news articles in our dataset. With the help of crowdsourcing, human contributors are prompted to extract specific and easy-to-extract information that helps match a given article to information representing "ground truth" obtained from the Syrian Violations Documentation Center. The information extracted is then used to cluster the articles into two separate sets using unsupervised machine learning. The result is a carefully annotated dataset consisting of 804 articles labeled as true or fake and that is ideal for training machine learning models to predict the credibility of news articles. Our dataset is publicly available at https://doi.org/10.5281/zenodo.2607278. Although our dataset is focused on the Syrian crisis, it can be used to train machine learning models to detect fake news in other related domains. Moreover, the framework we used to obtain the dataset is general enough to be used to build other fake news datasets around military conflicts, provided there is some corresponding ground-truth available.

Introduction

Well into its eighth year, the Syrian war continues to plunge into increasingly more troubling levels of violence. As world and regional powers get more embroiled in the conflict, serious questions arise surrounding the credibility of news documenting the facts of war in Syria. Unlike bias that is perceived in opinion columns, the spread of fake news surrounding the documentation of the war compromises not only the integrity of journalism, but can contribute to psychological warfare that drives the exodus and constant mobility of refugees, and hampers humanitarian planning for delivering aid to distraught communities. An evidence-based approach to combating fake news necessitates that one embark on a data scientific approach by which the general public can be assisted in automatically identifying fake news around the Syrian conflict with some reasonable assurance.

The lack of manually labeled fake news datasets around the Syrian conflict is the major bottleneck for advancing automatic fake news detection. In this work, we embark on the first step towards this goal, which necessitates acquiring and exploring media accounts from the Syrian war, and using them to generate labeled benchmark datasets. To this end, we develop a general-purpose distributed architecture leveraging some of the most recent technologies to handle Big Data such as Spark streaming, and Hbase, for pulling live-data streams and scraping for historical data from several media outlets. Using this news scraping framework, we explore a variety of media outlets representing mobilisation press, loyalist press, and diverse print media, and generate a representative corpus of these various types of media outlets. Our news corpus consists of *804* English news articles that report on war incidents that took place from 2011 to 2018.

Manually labeling news articles as true or fake is not only a difficult task; it can also be very subjective. This is particularly true for the case of news articles reporting on war incidents, where fake news might be accurately reporting a certain incident, and yet distorting some of the facts such as the number of casualties, the type of attack or the actor responsible for the attack. To avoid any subjectivity and obtain as accurate labels as possible, we employ a semi-supervised *fact-checking* labeling approach. More precisely, we tap on the database of the Syrian Violation Documentation Center (VDC)[1]. The VDC is a non-profit, non-governmental organization registered in Switzerland that documents human rights violations from the Syrian war. The VDC accepts funding solely from independent sources. Since its onset in

*This work is supported by the Collaborative Research Stimulus Fund (CRS) at the American University of Beirut

[1]https://vdc-sy.net/en/

2011, the VDC data records, in real time, war-related deaths as well as missing and detained people. As stipulated on its website, the VDC adheres to international standards for the documentation of its data.

The VDC database contains records about the casualties taking place throughout the Syrian conflict. Each record in the VDC database consists of "violations" information relating to the demographics, date, location, cause of death (e.g., type of weapon used), and status of the victim (civilian or non-civilian). To make use of the VDC database for labeling the news articles in our dataset, we first use crowdsourcing to extract information about casualties from our news articles. Note that extracting such information is considered an easy and objective task that does not require domain experts or access to information beyond those in the articles, compared to the original task of assessing the credibility of news articles. Once violations information have been extracted, we then map each news article to its closest VDC event and identify how accurately the article reports on the casualties compared to the truth from VDC. This can then be used as a cue to determine whether an article is fake or not. To this end, we cluster our dataset into two clusters based on how close they match the information in VDC and utilize the cluster centroids to determine the label of each cluster.

Using the above technique, our dataset consists of 426 true articles and 378 fake articles. Our dataset is balanced in the number of true vs fake articles, where 53% of our articles were labeled true and 47% were labeled fake. To validate the accuracy of our fact-checking labeling approach outlined above, a set of 50 articles were also manually assessed for credibility using a domain expert (one of the authors who is a researcher in media studies), who relied on the reputation of the source and the strength of attribution of news present in the article. We calculated the Cohen Kappa coefficient between the labels given by the domain expert and the labels obtained by our approach for these 50 articles, and the result was a Cohen Kappa coefficient of *0.43*. This weak agreement factor suggests that reputation-based classification and fact checking do not necessarily yield similar conclusions in the realm of fake news detection, particularly in the case of news articles reporting on war incidents, where fake news can seemingly appear credible except for the distortion of some facts, of which the domain expert might not be aware.

This paper is organized as follows. We start by reviewing related datasets and then give an overview of the VDC. We then describe how we scraped various media outlets to obtain the news articles in our dataset. We then describe how we annotated the articles to extract war violations information. Next, we describe how we used the articles' annotations to match them to VDC events and consequently label them as true or fake. We then provide some exploratory analysis of our dataset and finally present some future directions that shed light on the usability of our fake news dataset and the framework through which it was constructed.

Related Datasets

Most related fake news datasets focus on US political news, entertainment news or satire articles. Our goal is to build a dataset consisting of fake and true articles reporting on the Syrian war, which can be then be used to build robust machine learning models to automate the process of fake news detection in such context. We believe that relying on the available datasets for this task might be highly inadequate. News reporting on conflicts or wars are extremely unique. For instance, a common case of fake news surrounding the Syrian war involves inaccurate reporting on the type of an attack, the number of dead civilians or the actor responsible for a certain attack reported by the news articles. Nonetheless, we review related fake news datasets and contrast the way they were constructed with the methodology we undertook here to build our own dataset.

Shu et al. (Shu et al. 2018) build FakeNewsNet, a dataset of news labeled true and fake. In order to build this labeled dataset, they build crawlers that crawl fact-checking websites such as Politificat (for political news) and GossipCop (for entertainment news) to obtain news content for fake news and true news. Both of these websites provide analysis done by journalists and domain experts to label news articles as fake and real. The authors also crawl E! online for entertainment news pieces and consider all their news as real as they believe this source is a trusted source. FakeNewsNet however is restricted to the domains of US politics and entertainment news. The dataset is labeled by scraping news articles from fact-checking websites, which contains articles that are manually labeled as true or fake by journalists and domain experts. In addition, they assume that certain sources always provide true news. In our case, we are concerned with news articles surrounding the Syrian war, and there is no web sources we can tap on to retrieve manually labelled articles. Moreover, we do not label articles as fake or true based on whether we believe their source is trusted or not.

Torabi and Taboada (Torabi and Taboada 2018) introduce two datasets scraped from the web by leveraging links to news articles mentioned by fact-checking websites (Buzzfeed and Snopes) with their labels. Both datasets used by Torabi and Taboada are made up of full articles labeled by human experts. These labels were: true, mostly true, mixture of true and false, mostly false, and false stories out of this website. Again, both datasets focus on US politics or general international news. They do not contain any news articles reporting on conflicts or wars, which is the aim of our dataset construction. Moreover, the two datasets Torabi and Taboada were labeled manually as fake or true. In our case, we avoid this tedious manual labeling by making use of a reliable source, the Violation Documentation Center, which we use to fact check the claims in the news articles and to label them.

Golbeck et al. (Golbeck et al. 2018) built a dataset of fake news and satirical stories and restricted their dataset to American politics, recent articles, diverse sources, and no borderline cases. They identified a list of fake news and satirical websites and assigned researchers for each website to label each article scraped from the website as fake or satirical. Yang Wang (Yang Wang 2017) built the LIAR dataset, which includes 12.8K human labeled short statements from politifcat.com's API. They consider six fine-grained labels for each statement: pants-fire, false, barely-

	actor	cause_of_death	date_of_death	name	place_of_birth	place_of_death	demographic	status
0	Russian troops	Warplane shelling	2018-01-28	Batoul Hasan al-Haji	Kaferzita	Hama	Adult - Female	Civilian
1	Syrian government and affiliated militias	Shooting	2018-01-28	Qotaiba Jomea Kanj	Termala	Homs	Adult - Male	Non-Civilian
2	Syrian government and affiliated militias	Shooting	2018-01-28	Jihad al-Daikh	Telbeiseh	Homs	Adult - Male	Non-Civilian
3	Syrian government and affiliated militias	Shooting	2018-01-27	Halimeh Qasem al-Basha	Rastan	Homs	Adult - Female	Civilian
4	Syrian government and affiliated militias	Shooting	2018-01-26	Thamer al-Jamous	Waar	Homs	Adult - Male	Non-Civilian
5	Self administration forces	Shooting	2018-01-23	Khaled Odeh	Bab Sbaa	Homs	Adult - Male	Non-Civilian
6	Syrian government and affiliated militias	Shooting	2018-01-23	Alae al-Deen Naser	Halfaya	Hama	Adult - Male	Non-Civilian

Figure 1: Sample records in the VDC database

true, half-true, mostly-true, and true. These statements were sampled from news releases, TV/radio interviews, campaign speeches, tweets, etc.. The subjects of these tweets include economy, health-care, taxes, education, jobs, elections, etc. The LIAR dataset consists of labeled statements rather than full articles. In our approach, we work with full news articles instead. In addition, the LIAR datasets is again focused on statements related to US politics.

Rashkin et al. (Rashkin et al. 2017) published a collection of roughly 20k news articles from eight sources categorized into four classes: propaganda, satire, hoax and trusted. Again, they relied on the type of news sources to label each article, which is not applicable in our case. Finally, Rubin et al. (Rubin et al. 2016) published a dataset of 360 news articles. This dataset contains balanced numbers of individually evaluated satirical and legitimate texts. However, it focuses on detecting satire articles rather than fake news, which are reporting inaccurate information surrounding the Syrian war as in our case.

Violations Documentation Center (VDC)

The VDC is a non-profit, non-governmental organization registered in Switzerland that tracks and documents human rights violations from the Syrian war[2]. *The VDC accepts funding solely from independent sources*. Since its onset in 2011, the VDC data records, in real time, war-related deaths as well as missing and detained people. As stipulated on its website, the VDC *adheres to international standards for the documentation of its data*.

The VDC relies on reports from investigators and a ground network of internationally trained field reporters, who attempt to cover every governorate in Syria. Reporters collect data in three steps. First, initial information on one or more victims is gathered, from immediate and local sources (for example, hospitals, morgues, accounts of relatives/friends, etc.). Second, supporting information such as videos or photographs are sought. With this, the account gets confirmed and a record gets established. The last step consists in actively investigating key information originally missing around the reported violation. For each death, the record consists of information relating to the demographics, date, location, cause of death (e.g., type of weapon used), and status of the victim (civilian or non-civilian). The latter status corresponds to any combatant, be that a member of the government forces, opposition forces, or other armed factions. Data is available in both Arabic and English, despite that inconsistencies may occur between the two databases.

The VDC remains the *only human rights group documenting deaths in the Syrian conflict over the entire duration of the conflict*, and making the distinction between civilian or combatant status. It is also the *only one that endorses high risks in documenting the violations*. The VDC has been a source of valuable information for a wealth of notable public health publications on the human cost of the war in Syria (see (Fouad et al. 2017; Guha-Sapir 2018; Mowafi and Leaning 2018) for a few examples). It has been vetted and adopted by a large number of researchers working within the framework of the Lancet commission on Syria[3].

The VDC database records consist of the following fields:

- Name of causality
- Cause of death (e.g., shooting, shelling, chemical weapons, etc.)
- Gender and age group (i.e., adult male, adult female, child male, or child female)
- Type (civilian or non-civilian)
- Actor (e.g., rebel groups, Russian forces, ISIS, etc.)
- Place of death (e.g. Damascus, Hama, Aleppo, etc.)
- Date of death

Figure 1 displays a sample of some of the records in the VDC database.

Dataset Construction

In order to make sure to include articles reporting on as many major and controversial war events from the Syrian

[2]https://vdc-sy.net/en/

[3]www.thelancet.com/commissions/syria

Table 1: Major events in the Syrian war extracted from VDC

Event Date	VDC Peak	Peak Type	War Event
July 2015	ISIS	actor	Major Offensives against ISIS
July 2016	ISIS	actor	Major Offensives against ISIS
March 2016 - end of 2016	Russian	actor	Russian Attack on Syria
May 2016 - end of 2016	Syrian government	actor	Multiple Offensives All Over Syria
February 2015	warplane shelling	cause of death	Offensives against Kurds and Offensives against ISIS
July 2015	shooting	cause of death	Aleppo Offensive and Major Offensives against ISIS
July 2016	shelling	cause of death	Aleppo Offensive and Major Offensives against ISIS
March 2014 end of 2014	shooting	cause of death	Multiple Offensives All Over Syria
August 2013	chemical and toxic gases	cause of death	Ghouta Chemical Attack
August 2016	chemical and toxic gases	cause of death	Aleppo Chemical Attack
October 2017, November 2017	shooting	cause of death	The Raqqa Campain
April 2017	chemical and toxic gases	cause of death	Khan Sheikhoun Chemical Attack
July 2015	Aleppo	location	Aleppo Offensive
April 2017	Idlib	location	Khan Sheikhoun Chemical Attack
July 2016 August 2016	Aleppo	location	Aleppo Offensive
August 2013	Aleppo	location	Ghouta Chemical Attack

war as possible, we took a look at the peaks in the casualties reported in the VDC. These peaks were either peaks in a certain month (e.g. a sudden increase in the deaths by chemical weapons in Aleppo in August 2016), or long periods of similar events, but not necessarily peaks (e.g. deaths in Raqqa all over 2017 but increased in October and November, but not sudden peaks). Once we extracted these peaks, we researched the events that happened in Syria in the locations and dates of these peaks to find out the event that happened during that time. (E.g., the peak in chemical weapons in Aleppo marks the Aleppo chemical attack, and the peaks in Raqqa mark the Raqqa Campaign). Based on these observations from the VDC, we were able to extract some of the major events in the Syrian war as shown in Table 1.

Finally, we scraped various media outlets for the events described in Table 1. We used keywords relevant and specific to each of the events in order to make sure that we scrape all the articles reporting about this event. Using this approach, we were able to build a corpus of *804* news articles from the following set of sources: Reuters (libertarian), Etilaf (social responsibility press associated with the National Coalition for Syrian Revolution and Opposition Forces), SANA (mobilization press associated with the Syrian government), Al Arabiya (loyalist press associated with the government of the K.S.A.), Al Manar (a diverse print media outlet in Lebanon, associated with Hezbollah), Al Ahram (an Egyptian daily newspaper owned by the Egyptian government), Al Alam (an Arabic news channel broadcasting from Iran and owned by the state-owned media corporation Islamic Republic of Iran Broadcasting), Al Araby (a pan-Arab media outlet headquartered in London), Al Sharq Al Awsat (an Arabic international newspaper headquartered in London), Daily Sabah (Turkish pro-government daily published in Turkey), TRT (the national public broadcaster of Turkey), Jordan Times (an English daily newspaper based in Amman, Jordan), The Lebanese National News Agency (NNA), Sputnik (a Russian news agency established by the Russian government-owned news agency Rossiya Segodnya), and TASS(a major news agency in Russia).

Articles Annotation

In this section, we describe our approach to extract casualties information from news articles that can be checked against the VDC data. Recall that the VDC database contains records about the casualties taking place throughout the Syrian conflict.

We thus focus on extracting information about casualties from our news articles corpus as well. To be able to do this, we crowdsource the information extraction job using the crowdsourcing platform Figure Eight[4] (formally CrowdFlower). In particular, for each news article in our corpus, we ask *three* contributors (i.e., workers) on Figure Eight to answer the following questions:

1. What is the date (day, month and year) of the event reported in the article?
2. What is the location of the event reported?
3. How many civilian died in the event reported?
4. How many children died ?
5. How many women died?
6. How many non-civilians died?
7. Who does the article blame for the casualties?
8. How did the casualties die (cause of death)?

To avoid typing mistakes from contributors, we displayed possible answers for categorial questions using a drop-down menu. For questions that required reporting figures such as the number of casualties, we asked the contributors to insert the corresponding figures in free textboxes. For each question, we also included an "Article does not specify" option. For the date of the event, we use three drop down menus: one for day, one for month, and one for year. For the location of the event, the drop down menu includes the provinces in Syria that are listed in the VDC database.

In order to decide the payment of the contributors for each article they annotate, we split our articles into size categories

[4]https://www.figure-eight.com/

Table 2: Fleiss Kappa agreement of the contributors for each question

Question	Fleiss' Kappa Agreement
Number of Civilian Casualties	0.67
Number of Children Casualties	0.50
Number of Women Casualties	0.75
Number of Non-Civilian Casualties	0.56
Cause of Death	0.66
Actor	0.74
Place of Death Claim	0.51
Day	0.92
Month	1
Year	1

based on the number of words in each article, and set the prices of the contribution based on the article size.

To ensure high-quality annotations for our articles, we restricted the participation to only Level 3 contributors, who are a small group of contributors on Figure Eight with the most experience and the highest accuracy on past contributions. In addition, we made use of Figure Eight's test questions feature, where each contributor had to pass a quiz composed of articles from our dataset that were annotated by us. The answers that the contributors provided for these pre-annotated articles in quiz mode were then checked against our gold-standard answers. If a contributor passed the quiz mode, she was then allowed to participate in our job. Moreover, each page in our job consisted of five articles, one of which was a pre-annotated gold-standard article. Same as in the quiz mode, the answers that the contributors provided for the pre-annotated article were checked against our answers. These test questions were used to track the contributors performance on our job and were used to automatically remove contributors that have low-accuracy contributions. We set the minimum accuracy threshold of the job to 70%. This means that any contributor whose accuracy drops below this 70% threshold was automatically dropped from the job. Table 2 shows the Fleiss' Kappa agreement for the questions that we asked the contributors to answer. As can be seen from the table, we obtained moderate to perfect agreement among contributors on all questions (McHugh 2012).

We relied on a majority vote to pick for each article one answer per question. In particular, we used Figure Eight's aggregated report, which aggregates all of the responses for each individual article and returns the answer with the highest confidence for each question. In case an article does not contain an answer to any of our questions, we dropped this article from our dataset to ensure that all articles we have do report events that can be compared against the VDC data. Overall, we had *804* articles fully annotated using answers to all of our questions and *200* articles that did not have answers to any of our questions, which were dropped.

Finally, to ensure the validity of the annotations we obtained for the articles and to break ties in case there exists no majority vote, one of the co-authors (a graduate computer science student) reviewed all the annotations for every article in our corpus and corrected any mistakes in the aggregated annotations, and broke any ties.

Next, we display two example articles from our dataset along with the annotations we obtained for them.

- **Daily Sabah: Coalition airstrikes kill 85 civilians in Daesh-held villages in Syria's Manbij**

> July 19 2016. Airstrikes on Daesh-held villages in northern Syria killed at least 85 civilians on Tuesday as intense fighting was underway between the militants and U.S-backed fighters Syrian opposition activists and the extremist group said. Residents in the area blamed the U.S.-led coalition for the strikes that targeted two villages Tokhar and Hoshariyeh which are controlled by IS activists said. The villages are near the Daesh stronghold of Manbij a town that members of the PYD-dominated U.S.-backed Syria Democratic Forces (SDF) have been trying to capture in a weeks-long offensive. The Britain-based Syrian Observatory for Human Rights said at least 56 civilians including 11 children were killed in the strikes on the villages which also wounded dozens. Another activist group the Local Coordination Committees said dozens of civilians mostly families were killed. Turkeys official Anadolu Agency put the death toll at least at 85 adding that 50 civilians were also wounded in airstrikes. The Daesh-linked Aamaq news agency claimed 160 civilians mostly women and children were killed in Tokhar alone in a series of purportedly American airstrikes around dawn Tuesday. Postings on a Facebook page show images of people including children as they were being put in collective grave purportedly in the village of Tokhar. One photograph shows a man carrying the lifeless body of a child covered with dust while another shows a child partly covered by a blanket lying in a grave. Tuesdays casualties come on the heels of similar airstrikes on the Daesh-held town of Manbij on Monday when at least 15 civilians were reportedly killed. Meanwhile the headquarters of Daesh militants inside Manbij was captured as SDF forces pushed into the western part of the town over the weekend the U.S. military said in a statement on Tuesday. The headquarters which was located in a hospital was being used as a command center and logistics hub. The U.S.-backed Syrian rebels also took control of part of the town enabling civilians in the area to flee the fighting the statement said. The rebels were continuing to battle Daesh on four fronts for control of Manbij clearing territory as they pushed toward the center of the city the statement said. Daesh militants have staged counterattacks but the Syrian rebels have maintained momentum with the help of air strikes by the U.S.-led coalition the statement said. It said the coalition has carried out more than 450 air strikes around Manbij since the operation to take the town began. The U.S. Central Command said the coalition conducted 18 strikes on Monday and destroyed 13 Daesh fighting positions seven Daesh vehicles and two car bombs near Manbij. The Manbij area has seen intense battles between Daesh extremists and the Kurdish-led fighters who have been advancing under the cover of intense airstrikes by the U.S.-led coalition.

The following annotations were obtained for the shown article:

- Date of event: 19-07-2016
- Location of event: Manbij
- Actor: international coalition forces
- Cause of death: warplane shelling
- Number of civilian casualties: 85
- Number of children casualties: 11
- Number of women casualties: 0
- Number of non-civilian casualties: 0

- **SANA: Chemical Attack Kills Five Syrians in Aleppo**

> 03-08-2016. Chemical Attack Kills Five Syrians in Aleppo. At least five Syrians have been killed and a number of others injured in a chemical attack by foreign-sponsored Takfiri militants against a residential neighborhood in northwestern Syria. At least five Syrians have been killed and a number of others injured in a chemical attack by foreign-sponsored Takfiri militants against a residential neighborhood in northwestern Syria. Health director for Aleppo Mohammad Hazouri said five people died and eight others experienced breathing difficulties after artillery shells containing toxic gasses slammed into the Old City of Aleppo on Tuesday the official SANA news agency reported. Government sources said Takfiri terrorists had also used chemical munitions against civilians in the city of Saraqib in the Idlib province but militants accused government forces of carrying out the attack. Doctor Ibrahim al-Assad a neurologist in Saraqib said he treated 16 of 29 cases brought to his hospital on Monday night. He added that most of the victims were women and children and were suffering from breathing difficulties red eyes and wheezing. Rescuers and doctors in the city said the symptoms were similar to those caused by chlorine gas. The chemical raids come as the Syrian army is making progress in operations to retake Aleppo from militants who are seeing the noose tightening around them in the areas which they control.

The following annotations were obtained for the article shown above:

- Date of event: 03-08-2016
- Location of event: Aleppo
- Actor: unknown (claims terrorist organization but does not name the organization)
- Cause of death: chemical and toxic gases
- Number of civilian casualties: 5
- Number of children casualties: 0
- Number of women casualties: 0
- Number of non-civilian casualties: 0

Articles Labeling

Now that we have annotated our news article corpus, our next step is to match those articles against the VDC database in order to be able to deduce whether an article is fake or not. As explained in the previous section, we have extracted aggregated information about the war violations that took place in the events reported by the articles. On the other hand, the VDC database contains records about violations on an individual level. To be able to match facts from articles to those in the VDC database, we needed to aggregate the information in the VDC database. We achieved this by grouping the records in the VDC data by actor, date of death, cause of death, and place of death and then counted the number of children casualties, number of women casualties, number of civilian casualties, and number of non civilian casualties. Figure 2 shows a set of aggregated records from the VDC database. Each row in the figure can be viewed as one event, which could be matched against events reported in the news articles as we explain next.

Given an annotated news article, our goal is now to match it to an event from the aggregated VDC data. To do so, one might consider just selecting from the aggregated VDC data the event which has the same location, date, actor and cause of death as reported in the article. However, this beats the purpose of our work, as it assumes that all articles will report all such information accurately. In fact, many fake articles might actually report an event but blame another actor for the casualties in the event, or report a different cause of death (say denying that a chemical attack took place). For this reason, we only rely on the location of the event reported in the article and its date to match it to events in the aggregated VDC data. That is, given an article that reports an event that took place in location loc and on date d, we retrieve all the events from the aggregated VDC data where $place\_of\_death = loc$ and $d - w \leq date\_of\_death \leq d + w$ (i.e., within a window of w days). With this, we retrieve all the VDC events that took place in the same location and within a window of w days. The rationale behind this is that an article might report on an event that took place sometime in the past, or that the dates of death for certain people might only be confirmed in the VDC database a few days later.

Our event retrieval mechanism just described might result in retrieving zero or more events from the aggregated VDC data for any given event described in an article. This would typically be news articles that are either reporting on events that do are not recorded in the VDC database, or are distorting the location or time of the event. Figure 3 shows the average number of events retrieved per article for different window sizes. On the other hand, Figure 4 shows the percentage of articles that matched no VDC events at all for different window sizes. As can be seen from the two figures, the bigger the window size is, the more VDC events are matched for each article. Note that even for a window size of just 1 day, we only have about 10% of the articles with no matching VDC events, and on average less than 20 VDC events matched per article. We exclude those articles that do not match any VDC event, since we have no way of labeling them based on the semi-supervised learning technique we describe next.

Next, for each article, we extract the closest event from the aggregated VDC data that it matches, if any, based on the location and a date window. To do so,

	actor	place_of_death	date_of_death	cause_of_death	nb_women	nb_children	nb_civilians	nb_noncivilians
0	Al-Nusra Front	Aleppo	2014-08-11	Kidnapping - Execution	0	0	0	1
1	Al-Nusra Front	Aleppo	2014-08-13	Shooting	0	1	0	1
2	Al-Nusra Front	Aleppo	2014-10-11	Shooting	0	0	0	1
3	Al-Nusra Front	Aleppo	2015-01-03	Shooting	0	0	0	12
4	Al-Nusra Front	Aleppo	2015-01-28	Kidnapping - Execution	0	0	1	0
5	Al-Nusra Front	Aleppo	2015-05-08	Kidnapping - Execution	0	0	1	0
6	Al-Nusra Front	Aleppo	2015-05-16	Kidnapping - Execution	0	0	1	0
7	Al-Nusra Front	Aleppo	2015-07-31	Shooting	0	0	0	2
8	Al-Nusra Front	Aleppo	2016-01-22	Kidnapping - Execution	0	0	1	0
9	Al-Nusra Front	Aleppo	2017-04-09	Kidnapping - Torture	0	0	1	0

Figure 2: Sample aggregated records from the VDC database

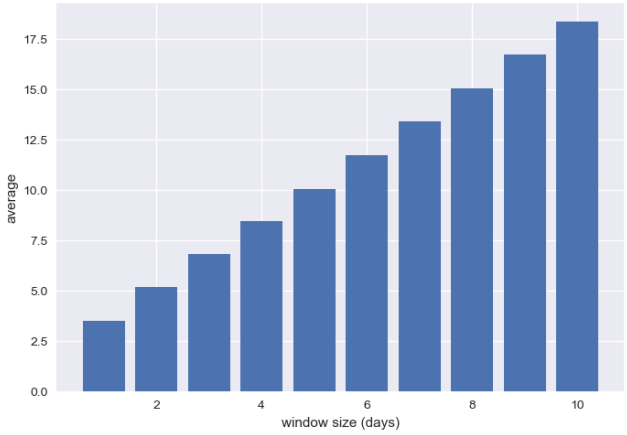

Figure 3: Histogram showing the average number of VDC events per article for different window sizes

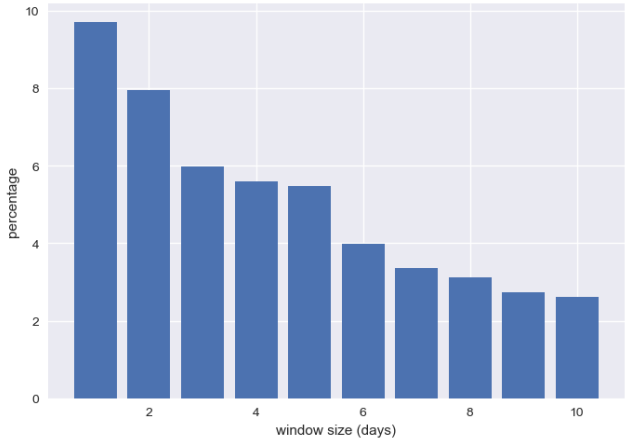

Figure 4: Histogram showing the percentage of articles that matched no VDC events for different window sizes

we rely on the *Gower's distance* between the VDC event and the article event using all the event features (i.e., $cause\_of\_death$, $actor$, $nb\_civilans$, $nb\_children$, nb_women and $nb\_noncivilians$). We use Gower's distance since our features consist of a mixture of numerical and categorial features (Huang 1998). Note that we do not include $place\_of\_death$ and $date\_of\_death$ as part of the feature space since those two attributes were used to initially select the candidate events to match against an article event. Once the closest VDC event has been retrieved, we use it to transform the article into a new feature space that represents how far the article's event is from the retrieved VDC event. The premise is that true news articles would be very close to their matched VDC events and fake ones will be far. To this end, we represent each news article using a 6-dimensional vector where the first two are binary features that represent whether or not the VDC event and the article's event agree on the cause of death and actor (1 if they both disagree and 0 if they agree). The rest are real-valued features that represent the difference between the number of women, children, civilian and non-civilian casualties reported in the article versus the closest VDC event. Particularly, given a type of violation, say civilian casualties, its corresponding feature $x\_civilians$ representing the difference in number of civilian casualties reported in the article versus VDC will be computed as follows:

$$x\_civilian = \frac{|nb\_civilians\_article - nb\_civilians\_vdc|}{nb\_civilians\_vdc}$$

where $nb\_civilians\_article$ is the number of civilian casualties as reported in the article and $nb\_civilians\_vdc$ is the number of civilian casualties as recorded in the VDC database. A similar formula was used to compute the rest of the real-valued features corresponding to the difference in number of casualties of women, children, and non-civilians.

We have mapped all the annotated articles in our dataset using the strategy highlighted above for different window sizes. We next devise a mechanism to classify the articles into true or fake that utilizes unsupervised machine learning. Given the articles in the new feature space for a given window size w, we cluster them into *two* clusters using K-prototypes clustering, since we have a mixture of categorial

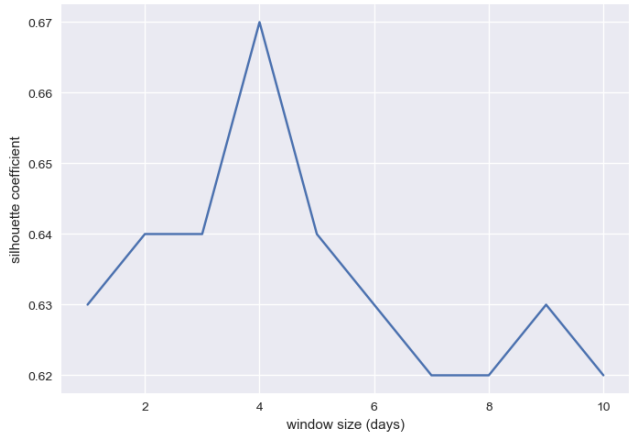

Figure 5: Average silhouette score of obtained clusters for various window sizes

and numerical features (Huang 1998). The intuition behind this is that the articles would cluster into two groups, one which contains articles that mostly coincide with the VDC data representing true articles and another corresponding to those articles that report facts that are not in line with those in their closest VDC event, corresponding to fake articles.

Figure 5 shows the average silhouette score of the obtained clusters for various window sizes. A higher average silhouette score indicates that each cluster contains items that are similar to each other and far from the items in the other cluster (Tan, Steinbach, and Kumar 2013). As can be seen from the figure, the window size that resulted in the highest average silhouette score is 4 days, with an average silhouette score of 0.67. It can also be noticed that as the window size increases, the silhouette score starts decreasing indicating poorer clustering of the articles. We thus pick the window size of 4 days as our final window size for which we base our credibility labeling of the articles on, which is a data-driven decision. Figure 6 shows the articles in the mapped feature space projected using PCA (Wold, Esbensen, and Geladi 1987) and the clusters they belong to for a window size of 4. As can be seen, there are some overlaps between the two clusters, and this can be attributed to two factors. First, the two-dimensional projection of the clusters might result in some distortion in the distances between the articles since our distances are based on six different dimensions. Second, the distance of articles to their matched VDC events is typically a spectrum, where some articles completely align with their VDC events in terms of all dimensions, others might only agree on some of these dimensions, and others might completely disagree with their VDC counterparts.

While we now have two clusters, these clusters remain unlabeled as to true or fake. Recall that each article was represented using a vector of six features, which corresponds to differences between the articles' claims and the VDC data on six attributes. The smaller these values are, the more consistent the article is with its corresponding VDC event. Since the articles in the true class are those that coincide the most

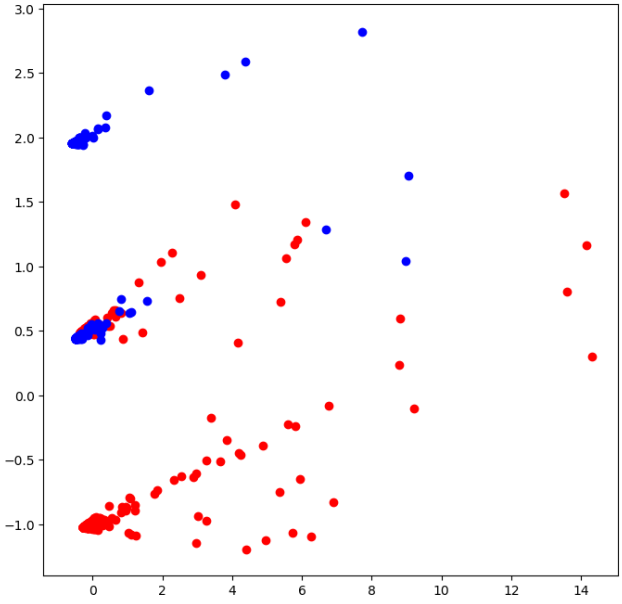

Figure 6: Two dimensional projection of the articles and their clustering for a window size of 4

with their corresponding VDC events, we label the cluster containing those articles with smaller feature values as the true class and the other cluster as the fake one. This was done by examining the centroids of both clusters, given by:

$$C\_true = [0.001, 0.002, 0.001, 0.001, 0.03, 0]$$

$$C\_fake = [0.03, 0.01, 0.01, 0.02, 0.5, 1]$$

where each of the vectors above are instances of the following feature vector: [$cause\_of\_death$, $actor$, $nb\_civilans$, $nb\_children$, nb_women, $nb\_noncivilians$].

As can be seen, the centroid of the cluster corresponding to true articles (C_true) has lower values with respect to all the features compared to the centroid of the fake articles (C_fake). Recall that we use 0 to represent agreement with VDC events and 1 to represent disagreement when it comes to our two binary features $cause\_of\_death$ and $actor$. Thus, a smaller value for those two features indicate more agreement with VDC data. Overall, using this method of labeling, we ended up with *426* true articles and *378* fake ones.

Next, we display the VDC events that are closest to our two example articles from the previous section and the labels we obtained for them using our approach.

- **SANA: Chemical Attack Kills Five Syrians in Aleppo**
  - location: Aleppo
  - number of civilians: 6
  - number of children, women, non-civilians: 0
  - actor: unknown
  - cause of death: shooting

  As can be seen, the closest VDC event to this article states that it is true that around five civilians were killed by an

Table 3: Example articles with disagreement between our approach and the media expert, and our labels

ID	Article Title	Big Scale Event	Our Label
1	At least 40 killed in Syrian weapons depot blast	Ghouta Chemical Attack	fake
2	ISIS recaptures Syrian gas fields kills 30 monitoring group says	Multiple Offensives All Over Syria	fake
3	Death toll in Syrian bombing raid on Aleppo rises to 76: monitor	Multiple Offensives All Over Syria	fake
4	80 Civilians Killed by Russian Airstrikes on Aleppo despite 48-Hour Truce	Russian Attack on Syria	fake
5	Iranian Militias Carry out Summary Executions against Civilians in Aleppo	Russian Attack on Syria	fake
6	Syrian Army Kills 11 ISIL Terrorists Destroy Their Posits in Deir Ezzor	Khan Sheikhoun Chemical Attack	true
7	Four people killed including a child and 25 others injured in terrorist attacks in Damascus	Russian Attack on Syria	true

Table 4: Agreement and disagreement of the example articles with VDC events

ID	Agreement with VDC	Disagreement with VDC
1	correct actor, nb children, and nb non-civilians	wrong cause of death, small difference in nb women, exaggerated nb civilians
2	correct cause of death, nb children, nb women, nb non-civilians	wrong actor, exaggerated nb civilians
3	correct cause of death, actor, nb non-civilians	exaggerated nb civilians, nb children, nb women
4	correct cause of death, actor, nb children, nb women	exaggerated nb civilians, small difference in nb non-civilians
5	correct cause of death, actor, nb children, nb civilians	small difference in nb women, understated nb civilians
6	correct cause of death, actor, nb civilians, nb children, nb women	small difference in nb non-civilians
7	correct cause of death, actor, nb non-civilians, nb women	small difference in nb civilians and nb children

unknown organization in Aleppo. However, these civilians were killed by shooting and not by chemical and toxic gases. This article was labeled *fake* by our method.

- **Daily Sabah: Coalition airstrikes kill 85 civilians in Daesh-held villages in Syria's Manbij**
  - location: Manbij
  - number of civilians: 69
  - number of children: 9
  - number of women and non-civilians: 0
  - actor: international coalition forces
  - cause of death: warplane shelling

  As can be seen, the closest VDC event to this article also states that it is true that around 85 civilians were killed by the international coalition's warplane shelling in Aleppo. The article has the correct actor, cause of death, an almost correct number of children and a very close number of civilians. This article was labeled *true* by our method.

Finally, to validate the accuracy of our fact-checking labeling approach, a media studies expert (a co-author of the paper) undertook her own manual labeling of 50 articles using techniques anchored around the reputation of the source and the strength of attribution of news present in the articles, independently of the VDC. We then measured agreement between the labels obtained by our approach and that of the domain expert. The Cohen Kappa coefficient between our labels and the media expert's labels for these 50 articles was 0.43.

Out of the fifty labels provided by our media expert, only fourteen disagreed with our labels (i.e., 28%). In Table 3, we show 7 examples out of those 14, along with the labels obtained using our approach. Those examples were the outliers with the highest disagreement with the VDC (rows 1-5) and the lowest disagreement with the VDC (rows 6 and 7). In Table 4, we show how these articles agreed and disagreed with the closest VDC event in terms of our six features (cause of death, actor, numbers of civilians, children, women, and non-civilians). As can be observed from the table, many fake articles might diverge only by denying that a certain type of attack took place, by blaming a different actor than the actual one responsible for the incident or attack, or by overstating or understating the number of casualties. On the other hand, some articles might not have a strong source attribution or emerge from a high-reputation sources. Nonetheless, they might still indeed be true. This highlights the difficulty of manually assessing the credibility of news articles reporting on war incidents. It also suggests that reputation-based classification and fact checking do not necessarily yield similar conclusions in the realm of fake news detection, particularly in the case of news articles reporting on war incidents, where fake news can seemingly appear credible, except for the distortion of some facts, of which the domain expert might not be aware.

Exploratory Analysis

In this section, we perform some exploratory analysis of FA-KES, our fake news dataset around the Syrian war. Recall that FA-KES consisted of a total of 804 news articles, of which 426 were labeled true ($\approx 53\%$) and 378 were labeled fake ($\approx 47\%$).

Exploring the the number of articles labeled fake per month during the Syrian war, we notice that the dates with the *peaks* of fake articles from our dataset were during the following events: April 2017, when the Khan Sheikhoun chemical attack took place, summer 2016, during which the Aleppo offensive and other major offensives against ISIS took place, August and September 2016, around the time of the Aleppo chemical attack, and August and September 2013, during the times of the Ghouta chemical attack. We also studied the distribution of articles that were labeled true/fake for each news source category (i.e., news sources that are pro (Syrian) regime, those against the regime as well as neutral ones). We observed that *over 70%* of the pro-regime news articles were labeled as fake in our dataset, compared to *less than 30%* for against-regime articles, and

around 50% for neutral ones. Judging from the peaks around which fakes news in our dataset have been reported, this might be attributed to a desire to deny that certain war crimes have been committed by regime forces/coalition at the alleged times or places.

Conclusion, Limitations and Perspectives

To the best of our knowledge, we have produced the first dataset in the literature that presents fake news surrounding the conflict in Syria. Our work is attained using a general framework that can be easily extended to other controversial events being reported on using conflicting accounts, provided some ground truth is available and generated by "witnesses". Our approach is data-driven rather than model-driven, providing for a fact-checking fake news labeling mechanism with the help of crowdsourcing and unsupervised learning. It is also carefully software-engineered to make us of Big Data platforms, allowing the tool to scale as much as needed. Our dataset can be readily used to train supervised machine learning algorithms to detect fake news automatically without the need for "ground truth" data. This will permit the automatic fake news detection mechanism by the general public. Our dataset is publicly available at https://doi.org/10.5281/zenodo.2607278[5]. Although our dataset is focused on the Syrian crisis, it can be used to train machine learning models to detect fake news in other related domains. Moreover, the framework we used to obtain the dataset is general enough to be used to build other fake news datasets around military conflicts, provided there is some corresponding ground-truth available. Current limitations that require further investigation are related to the poor agreement with the labels provided for a small subset of the dataset by a media studies expert relying on reputation-based analysis, and to the disagreement that surfaced between the annotations provided by the crowd workers and the corresponding annotations from the VDC. In future work, we plan to validate this by building more fake news datasets in other domains using our framework. We will also attempt to develop an information extraction approach to automatically extract war violations information from news articles that can be then matched against VDC data or other ground truth databases. Finally, we have already built a fully-supervised machine-learning models to automatically detect fake news using high level signals pertinent to the Syrian military conflict (e.g. sectarian tone, consistency with respect to the VDC), and tested these models on news articles related to the Syrian war as well as other fake news datasets. This work is currently in progress.

Acknowledgement

We thank the American University of Beirut for funding this work through the Collaborative Research Stimulus.

[5]The dataset is available for research purposes only, and any commercial use is forbidden. All publications using this dataset have to acknowledge this by citing the present article.

References

Fouad, F. M.; Sparrow, A.; Tarakji, A.; Alameddine, M.; El-Jardali, F.; Coutts, A. P.; Arnaout, N. E.; Karroum, L. B.; Jawad, M.; Roborgh, S.; Abbara, A.; Alhalabi, F.; AlMasri, I.; and Jabbour, S. 2017. Health workers and the weaponisation of health care in Syria: a preliminary inquiry for The Lancet–American University of Beirut Commission on Syria. *The Lancet* 390(10111):2516–2526.

Golbeck, J.; Mauriello, M.; Auxier, B.; Bhanushali, K. H.; Bonk, C.; Bouzaghrane, M. A.; Buntain, C.; Chanduka, R.; Cheakalos, P.; Everett, J. B.; Falak, W.; Gieringer, C.; Graney, J.; Hoffman, K. M.; Huth, L.; Ma, Z.; Jha, M.; Khan, M.; Kori, V.; Lewis, E.; Mirano, G.; Mohn IV, W. T.; Mussenden, S.; Nelson, T. M.; Mcwillie, S.; Pant, A.; Shetye, P.; Shrestha, R.; Steinheimer, A.; Subramanian, A.; and Visnansky, G. 2018. Fake news vs satire: A dataset and analysis. In *Proceedings of the 10th ACM Conference on Web Science*.

Guha-Sapir, D. 2018. Patterns of civilian and child deaths due to war-related violence in Syria: a comparative analysis from the Violation Documentation Center dataset, 2011–16. *The Lancet Global Health* 6(1).

Huang, Z. 1998. Extensions to the k-means algorithm for clustering large data sets with categorical values. *Data mining and knowledge discovery* 2(3):283–304.

McHugh, M. L. 2012. Interrater reliability: the kappa statistic. *Biochemia medica: Biochemia medica* 22(3):276–282.

Mowafi, H., and Leaning, J. 2018. Documenting deaths in the Syrian war. *The Lancet Global Health* 6(1).

Rashkin, H.; Choi, E.; Jang, J. Y.; Volkova, S.; and Choi, Y. 2017. Truth of varying shades: Analyzing language in fake news and political fact-checking. In *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing*, 2931–2937.

Rubin, V.; Conroy, N.; Chen, Y.; and Cornwell, S. 2016. Fake news or truth? using satirical cues to detect potentially misleading news. In *Proceedings of the Second Workshop on Computational Approaches to Deception Detection*, 7–17.

Shu, K.; Mahudeswaran, D.; Wang, S.; Lee, D.; and Liu, H. 2018. Fakenewsnet: A data repository with news content, social context and dynamic information for studying fake news on social media. *arXiv preprint arXiv:1809.01286*.

Tan, P.-N.; Steinbach, M.; and Kumar, V. 2013. Data mining cluster analysis: basic concepts and algorithms. *Introduction to data mining*.

Torabi, F., and Taboada, M. 2018. The data challenge in misinformation detection : Source reputation vs . content veracity.

Wold, S.; Esbensen, K.; and Geladi, P. 1987. Principal component analysis. *Chemometrics and intelligent laboratory systems* 2(1-3):37–52.

Yang Wang, W. 2017. Liar, liar pants on fire: A new benchmark dataset for fake news detection. In *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics, ACL 2017*, 422–426.