UDC 621.391

H. Pievtsov, A. Feklistov

Ivan Kozhedub Kharkiv National Air Force University, Kharkiv

THE ANALYSIS OF TERRORIST ATTACKS IN UKRAINE
BASED ON OPEN-SOURCE DATA SET

The article considers results of exploratory data analysis of terrorist attacks in Ukraine based on open-source data set "The Global Terrorism Database" made by National Consortium for the Study of Terrorism and Responses to Terrorism. The original data set has information about 1583 terrorist attacks in Ukraine and reflects changes in the security situation in 2014-2015 years. The outcomes of research can be used for the planning of counterterrorism and information operations. These ranges from basic information about what types of threats and tactics are prevalent to insights on what types of counterterrorism strategies are most likely to be effective.

Keywords: information operations, terrorism, open-source data set, exploratory data analysis.

Introduction

Formulation of the problem in general. The recent changes in security and defence situation in Ukraine concerning Crimea occupation and Anti-Terrorist Operation (ATO) are key elements for a new challenge to Ukranian Armed Forces (UAF) known as a terrorism attacks (TA). It defined as a "threatened or actual use of illegal force and violence by a non-state actor to attain a political, economic, religious, or social goal through fear, coercion, or intimidation" [1–9].

One of the approaches for analysis TAs in Ukraine is using "The Global Terrorism Database (GTD) made by National Consortium for the Study of Terrorism and Responses to Terrorism (START) in an effort to increase understanding of terrorist violence so that it can be more readily studied and defeated. The GTS is an open-source database including information on terrorist events around the world from 1970 through 2015 with more than 150,000 cases [6].

For each case it includes information on at least 45 features, with more recent incidents including information on more than 120 features. In general, it includes information on more than 75,000 bombings, 17,000 assassinations, and 9,000 kidnappings since 1970 [6].

The GTD has been leveraged extensively in scholarly publications, reports, and media articles, and the academic community regulary conduct researches of the data set for international and domestics challenges. For example, there are researches for TA in United States and Nepal [8–9].

In the same time the analsyis of existed publications shown that for today there isn't any known research of TA in Ukraine based on this data set.

The purpose of the article is to present results of exploratory data analysis of TA in Ukraine based on open-source GTD data set.

Statement of Materials Research

The information contained in the GDB is based on reports from a variety of open media sources. The GTD unlike many other event databases includes systematic data on domestic as well as transnational and international terrorist incidents that have occurred during this time period. It's based entirely on publicly available, unclassified source materials. These include media articles and electronic news archives, and to a lesser extent, existing data sets, secondary source materials such as books, journals, and legal documents.

With the expansion of online media, the GTD developers created a "hybrid" data collection strategy. It leverage automated processes (natural language processing, machine learning models) to sift through millions of news articles each month. Over 4,000,000 news articles and 25,000 news sources on any topic published daily worldwide were reviewed to collect incident data from 1998 to 2015 alone. Each month, GTD researchers at START review approximately 16,000 articles and identify attacks to be added to the GTD. Currently it's the most comprehensive unclassified data base on terrorist events in the world [5].

The GTD Codebook "Inclusion Criteria and Variables" outlines the features that constitute the GTD and defines their possible values. There are 9 groups of features: GTD ID and date, incident information, incident location, attack information, weapon information, target/victim information, perpetrator information, casualties and consequences, and additional information and sources.

The main analytical insights extracted from these groups of features are the following.

1. GTD ID and date. The GTD that was used in research was downloaded from the website www.kaggle.com and has 156,772 observations and 137

features for 206 countries. There are 1,583 or 1% observations for Ukraine including 491 unique dates of incidents for 19 years of observations. For the first 16 years the number of attacks is less than 10. The situation dramatically changed in 2014 year in which were 891 incidents, and in the next 2015 year with 637 incidents. So, the Ukraine has 1528 or 97% of all attacks in 2014-2015 years.

2. Incident information. There are 1580 and 1581 cases in which an incident met inclusion criteria "Political, economic, religiuos, or social goal" and "Intention to coerce, intimidate or publicize to larger audiences", and 797 cases in which an incident met and in 786 not met inclusion criteria "Outside International Humanitarian Law".

In the 375 cases an attack was a part of "multiple" incident, 96 cases in which the duration of an incident extended more than 24 hours, and 373 cases when an attack is part of a coordinated, multi-part incident.

3. Incident location. There are 35 first order subnational administrative regions and 330 names of city, village, or town in which the incident occurred. The city with biggest summary number of attacks is "Donetsk" (tabl. 1).

Table 1

Top 10 cities by the number of attacks

№	City	Attacks	Percent (%)
1.	Donetsk	185	19.21
2.	Shyrokyne	52	5.40
3.	Luhansk	47	4.88
4.	Shchastya	46	4.78
5.	Slovyansk	46	4.78
6.	Pisky	45	4.67
7.	Kiev	43	4.46
8.	Odessa	41	4.25
9.	Avdiivka	40	4.15
10.	Debaltseve	34	3.53

There are 462 cases in which an incident occurred in the immediate vicinity of the city and in 164 cases an incident has additional information about the location (for example, "The incident occurred near the Donetsk Airport").

4. Attack information. Two features define the general method of attack and often reflect the broad class of tactics used (tabl. 2).

Table 2

Top 5 methods of attacks

№	Method of attacks	Sum	Percent (%)
1.	Bombing/Explosion	886	55.97
2.	Armed Assault	392	24.76
3.	Hostage Taking (Kidnapping)	104	6.57
4.	Facility/Infrastructure Attack	89	5.62
5.	Assassination	30	1.89

The 1423 of attacks are classified as a "successful" according to the tangible effects. And, there are only 2 cases in which perpetrator did not intend to escape from the attack alive.

5. Weapon information. The most popular general type of weapon and weapon sub-types used in the incident are given in the tabl. 3.

Table 3

Most popular general type of weapon and weapon sub-types used in the incident

№	General type of weapon and weapon sub-types	Sum	Percent (%)
1. Weapon types			
1.1.	Explosives/Bombs/Dynamite	1088	68.73
1.2.	Firearms	282	17.81
2. Sub-weapon types			
2.1.	Projectile (rockets, mortars, RPGs)	830	52.43
2.2.	Unknown Gun Type	230	14.53
2.3.	Unknown Explosive Type	140	8.84

The additional information described conventions follow "Second Weapon Type and Sub-Type" with the popular value "Firearms", "Unknown Gun Type", "Automatic Weapon" and "Rifle/Shotgun (non-automatic)".

There are 350 unique details with any pertinent information on the type of weapon(s) used in the incident (for example, "Artillery was used in the attack").

6. Target/victim Information. The most popular general type and sub-types of target/victim are given in tabl. 4.

Table 4

The most popular general type and sub-types of target/victim in incidents

№	General type and sub-types of target/victims	Sum	Percent (%)
1. General types of target/victims			
1.1.	Military	840	53.06
1.2.	Private Citizens & Property	287	18.13
2. Sub types of target/victims			
2.1.	Military Personnel (soldiers, troops, officers, forces)	459	28.99
2.2.	Military Checkpoint	150	9.48
2.3.	Village/City/Town/Suburb	148	9.35
2.4.	Military Barracks/Base/Headquarters/Checkpost	111	7.01

The main entities from defence and security sectors are "Armed Forces of Ukraine", "State Border Guard Service of Ukraine", "Militsiya", and "National Guard of Ukraine".

There are 450 specific target/victim like that was targeted and/or victimized and is a part of the entity named above. The most popular target/victim are "Soldiers", "Town", "Anti-Terrorist Operation Soldiers", "Checkpoint", and "Donetsk Sergey Prokofiev International Airport", and for the second target/victim type "Private Citizens & Property", "Journalists & Media", and "Utilities".

There were 19 nationalities of the target that was attacked, and the main are from "Ukraine", "International", "Russia", and "United States".

7. Perpetrator information. There are 16 Perpetrator's Group Names as names of the group that carried out the attack, and "Perpetrator Sub-Group Name" with additional qualifiers or details about the name of these groups.
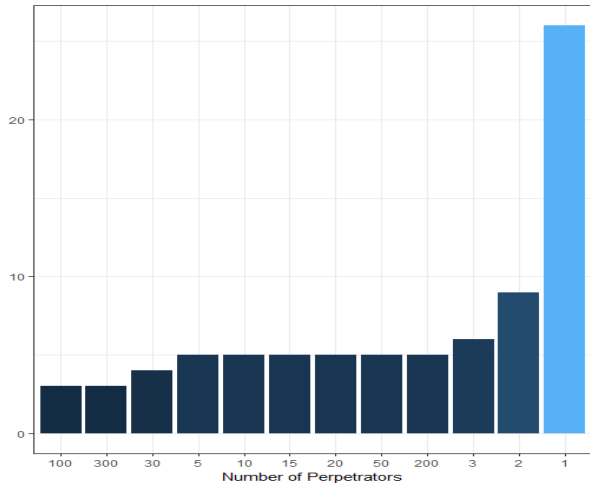

Fig 1. The total number of terrorists participating in the incidents

The total number of terrorists participating in the incidents is given on the fig 1. The most popular values are 1-3 perpetrators with "1" at the top.

The most popular are "Donetsk People's Republic (600 or 37.90 %), "Luhansk People's Republic" (206 or 13.01 %), and "Donbass People's Militia" (18 or 1.14 %). In 1564 cases the perpetrator attribution(s) for the incident are not suspected.

In most known cases a group or person claimed responsibility for the attack (94 or 5.94%). The modes for claim of responsibility used by claimants might be useful to verify authenticity and track trends in their behavior.

There are 71 motive for the attacks like "The specific motive is unknown; however, Ukrainian officials speculated that the bridge was destroyed in order to halt the advances of Ukrainian military forces in the region".

8. Casualties and consequences. The total number of fatalities as a number of total confirmed fatalities for the incident that includes all victims and attackers who died as a direct result of the incident are "0" (1068 or 67.47 %) and "1" (157 or 9.92%) in most cases except 3 incidents were killed more than 100 people (298, 201, and 143).

The total number of injured as a number of confirmed non-fatal injuries to both perpetrators and victims are "0" (963 or 60.83 %) and "1" (135 or 8.53%) in most cases except 3 incidents were injured more than 100 people (157 and 140).

The number of incidents resulted in property damage is equal ("yes" – 505 and "no" – 498 cases). In 308 known cases the extent of the property damage is defined as "Minor (likely < $1 million)".

There are 326 specific details about the property that was damaged in an attack, such as the type of vehicle that was destroyed, the areas or parts of a building that were damaged, or the types of assets that were stolen. For example, there are 26 cases where "A building was damaged in this attack".

One or two victims were taken hostage or kidnapped, and they spent are between 1-8 days in this state in 111 cases in total.

In 2 cases the country that Kidnappers/Hijackers diverted to is "Russia". In one case the incident involved a demand of monetary ransom. Also, there are 9 ransom notes like "The assailants demanded the release of imprisoned pro-Russian separatists in exchange for the safe return of the hostages".

The most popular value of eventual fate of hostages and kidnap victims is "Hostage(s) released by perpetrators" (57 or 3.60%).

Feature selection. The data set contains many features some of them are redundant and can be removed without incurring much loss of information for predictive modeling. The analysis of features clustering in correlation matrix based on using "network_plot()" function from R package "corrr" that allows exploring correlations through visualisation [10].

The plot shows a point for each feature rather than for each correlation. The proximity of the features to each other represents the overall magnitude of their correlations. Each path represents a correlation between the two features that it joins. The color, width and transparency of the line represent the strength of the correlation (fig. 3, tabl. 5).
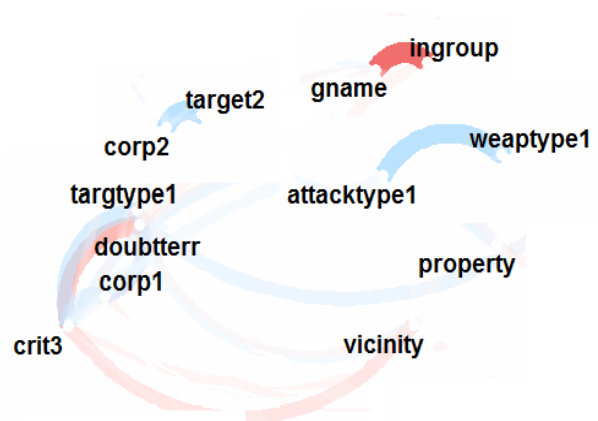

Fig. 2. Visualization of correlations between features by using "network_plot()" function from "corrr" R package

Table 5

Text description correlations between features

№	Description	Correla-tion
	Positive correlations	
1.	The general method of attack (attacktype1) with general type of weapon used in the incident (weaptype1).	0.76
2.	Second name of corporate entity or government agency (corp2) that was targeted or victimized) with second specific target (victim) name (target2)	0.76
3.	Third inclusion criteria (crti3) with the general type of target (victim) (targtype1)	0.57
	Negative correlations	
4.	The name of the group that carried out the attack (gname) with the number of terrorists in group during the incident (ingroup)	-0.97
5.	Third inclusion criteria (crti3) with a feature that reflect doubt as to whether the incident is an act of terrorism (doubtterr)	-0.66

Fig. 3. Visualiation of "cluster" features extraction

The research was done by using the following R packages "tidyverse", "data.table", "feather", "cluster", "corrplot", "shiny", "ggplot2" [10–12].

Conclusions

The analysis of GDT open-source data set is a new opportunity for improvement Ukrainian security and defence decision support information systems. The outcomes of research can be used for the planning of counterterrorism and information operations.

These ranges from providing fairly basic information about what types of threats and tactics are prevalent across various jurisdictions, and how they vary over time to more sophisticated analyses that attempt to provide insights on what types of counterterrorism strategies are most likely to be effective in a given context.

The main perspective ways for further research are advanced data analyses of GTD features group by cities, targets, regions, perpetrators, attacks, weapons, including natural language processing text features, and develop predictive models based on this data set.

Feature extraction. As a part of feature engineering process, new features were extracted from existed ones. The most popular partitioning clustering K-means method was used to identify new "cluster" features. The optimal number of clusters is 4 and was defined by using Elbow method, Bayesian Inference Criterion for k means, optimum average silhouette width criteria and Calinski criteria for an optimal number of clusters (fig. 3).

The selected and extracted features are the basis for further predictive modeling of different analytical insights and predictions.
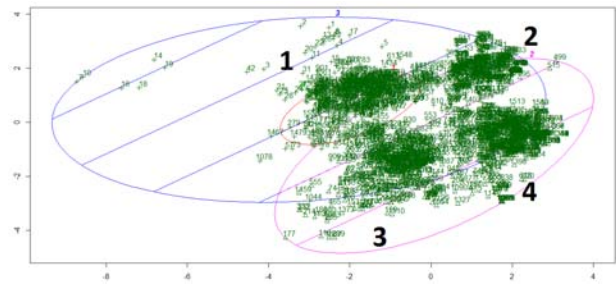
References

1. Alimpiev, A.M. (2015), "*Suchasne ozbroyennya i viys'kova tekhnika Zbroynykh Syl Rosiys'koyi Federatsiyi. Dovidnyk uchasnyka ATO*" [*The Guide of ATO participant. Arms and military equipment of the Russian Federation Armed Forces*], Kharkiv, 732 p.

2. Pievtsov, H.V., Hordiienko, A.M., Zalkin, S.V., Sidchenko, S.O., Feklistov, A.O. and Khudarkovskyi, K.I. (2017), "*Informatsiino-psykholohichna borotba u voiennii sferi: monohrafiia*" [*The information and psychological struggle in the military sphere*], Rozhko S.H., Kharkiv, 276 p.

3. Pevtsov, G.V., Zalkin, S.V., Sidchenko, S.O. and Khudarkovsky, K.I. (2013), "*Informatsiyna bezpeka u voyenniy sferi: problemy, metodolohiya, systema zabezpechennya: monohrafiya*" [*Information security in the military sphere: problems, methodology, system of provision: monographiia*], Digital printing-house No. 1, Kharkiv, 272 p.

4. Pevtsov, H.V., Gordienko, A.M., Zalkin. S.V., Sidchenko, S.O. and Khudarkovskii, K.I. (2015), "Dosvid i kontseptsiyi vedennya informatsiynoyi borot′by u providnykh krayinakh svitu" [Experience and Concepts of Information Fighting in the Leading Countries of the World], *Science and Technology of the Air Force of Ukraine*, No. 1 (18), pp. 12-16.

5. Alimpiev, A.M. and Pevtsov, G.V. (2017), "Osoblyvosti hibrydnoyi viyny Rosiys′koyi Federatsiyi proty Ukrayiny. Dosvid, shcho otrymanyy Povitryanymy Sylamy Zbroynykh Syl Ukrayiny" [The features of Russian hybrid war against Ukraine. Experience gained by the Air Forces of the Armed Forces of Ukraine], *Science and Technology of the Air Force of Ukraine*, No. 2 (27), pp. 19-25.

6. *Information on more than 150,000 Terrorist Attacks and Codebook "Inclusion Criteria and Variables*", National Consortium for the Study of Terrorism and Responses to Terrorism (START), www.start.umd.edu/gtd (accessed 12 July 2017).

7. LaFree, G., Dugan, L., Fogg, H.V. and Scott, J. (2006), *Building a Global Terrorism Database*, Diane Publishing Co., 209 p.

8. Abigail, L. (2017), *Terrorist Attacts in US*, www.bit.ly/2tI5H8W (accessed 20 Jule 2017).

9. Scottish, C. (2017), *Terrorism in Nepal 1970-2015*, www.bit.ly/2rLvTDd (accessed 12 Jule 2017).

10. Simon, J. (2016), *Correlation network_plot() with corrr*, www.bit.ly/2sXqZzJ (accessed 28 July, 2016).

11. Kabacoff, R. (2015), *R in Action: Data Analysis and Graphics with R (2nd Edition),* Manning Publications Co, 580 p.

12. Grolemund, G. and Wickham, H. (2017), *R for Data Science. Import, Tidy, Transform, Visualize, and Model Data,* O'Reilly Media, Inc., 520 p.

Відомості про авторів:

Пєвцов Геннадій Володимирович
доктор технічних наук професор
заступник начальника Харківського
національного університету Повітряних Сил
ім. І. Кожедуба з наукової роботи,
Харків, Україна
https://orcid.org/0000-0002-0426-6768
e-mail: pgvn@i.ua

Феклістов Андрій Олександрович
кандидат технічних наук
старший науковий співробітник
провідний науковий співробітник Харківського
національного університету
Повітряних Сил ім. І. Кожедуба,
Харків, Україна
https://orcid.org/0000-0003-1686-8333
e-mail: afek@ukr.net

Information about the authors:

Pievtsov Hennady
Doctor of Technical Sciences Professor
Deputy Head of Ivan Kozhedub Kharkiv National
Air Force University in Science,
Kharkiv, Ukraine
https://orcid.org/0000-0002-0426-6768
e-mail: pgvn@i.ua

Feklistov Andrey
Candidate of Technical Sciences Senior Research
Leading Researcher of Ivan Kozhedub
Kharkiv National Air Force University,
Kharkiv, Ukraine
https://orcid.org/0000-0003-1686-8333
e-mail: afek@ukr.net

АНАЛІЗ ТЕРОРИСТИЧНИХ АТАК В УКРАЇНІ НА ОСНОВІ ВІДКРИТИХ ДАНИХ

Г.В. Пєвцов, А.О. Феклістов

У статті розглядаються результати розвідувального аналізу даних щодо терористичних атак в Україні на основі відкритого набору даних "База даних глобальної тероризму" Національного консорціуму з вивчення та реагування на тероризм. Оригінальний набір даних містить інформацію про 1583 терористичні атаки в Україні та відображає зміни в ситуації безпеки в 2014-2015 роках. Результати досліджень можуть бути використані для подальшого прогнозного моделювання та планування антитерористичних та інформаційних операцій. Вони можуть бути представлені від базової інформації про типи загроз і тактики їх застосування до уявлень про ефективні стратегії боротьби з тероризмом в заданому контексті.

Ключові слова: інформаційна операція, тероризм, відкритий набір даних, розвідувальний аналіз даних.

АНАЛИЗ ТЕРРОРИСТИЧЕСКИХ АТАК В УКРАИНЕ НА ОСНОВЕ ОТКРЫТЫХ ДАННЫХ

Г.В. Певцов, А.А. Феклистов

В статье рассматриваются результаты разведочного анализа данных о террористических атаках в Украине на основе открытого набора данных «Глобальная база данных терроризма», подготовленного Национальным консорциумом по изучению и реагированию на терроризм. Исходный набор данных содержит информацию о 1583 терростических актах в Украине и отражает изменения относительно безопасности в 2014-2015 годах. Результаты исследования могут быть использованы для дальнейшего прогнозного моделирования и планирования контртеррористических и информационных операций. Они могут быть представлены от базовой информации о типах угроз и тактики их применения до представлений об эффективных стратегиях борьбы с терроризмом в заданном контексте.

Ключевые слова: информационная операция, терроризм, открытый набор данных, разведывательный анализ данных.