SATP Terrorist Activity Annotation Codebook

Javier Osorio (University of Arizona)
Shahreen Psyche (University of Texas - Dallas)

July 27, 2020

Contents

1 Introduction

The purpose of this exercise is to generate human annotations of terrorist activity from the South Asia Terrorism Portal (SATP) between 2010 and 2020. These annotations will serve as input for a Machine Learning model and as Gold Standard Records for validation purposes. To code terrorist activity, we will use the guidelines of the Global Terrorism Database (GTD) project.

2 Terrorist Activity

The South Asia Terrorism Portal (SATP) website (https://satp.org/terrorist-activity/) contains a timeline per country displaying a collection of brief narratives summarizing in-

cidents of terrorist activity per country. Figure 1 below shows an example of the timeline narratives of India.

Figure 1: SATP timeline of terrorist activity in India

The SATP narratives provide a description of incidents broadly related to terrorism. However, not all those narratives are relevant for the purposes of our research. Since we are primarily focused on terrorist activity, we need to take a closer look at the SATP narratives to identify which incidents are relevant and which are not. Then, we need to classify some specific information about those events that are indeed relevant to our study.

To guide our research, we follow previous studies on terrorism. In particular, we use as guideline the codebook of the Global Terrorism Database (GTD) project `https://start.umd.edu/gtd/downloads/Codebook.pdf` to identify the main characteristics of terrorist activity.

2.1 Event Data

Before getting into the details of terrorist activity, it is important for us to define what constitutes an event and its main components.

We define **event data** as a categorical description of a *source* actor conducting an *action* against a *target* actor in a given *location* at a certain *time*. In this way, an event provides information about someone doing something to someone else in a location and a given date.

The core elements of an event are:

- **Source**: Is the actor who initiates the action. For the purposes of this research, we consider only actions initiated by terrorist organizations. In this way, a terrorist group should always be the source actor in this research.

- **Action**: Indicates the action being conducted. More details about this in section 2.2.

- **Target**: It refers to the actors targeted by the terrorist activity. More details about this in section .

- **Location**: Depending on the level of information, we could classify the location as country, state, or county.

- **Time**: It is the date when the action took place.

Let's see an example to understand the core event elements in a sample SATP data:

```
Unidentified militants killed a local journalist, Sultan Mahmoud Khairkhwah,
in in the west of Khost city on March 16, reports Tolo News.  The slain journalist
was working for Zan TV network in Khost province.
```

- **Source**: Unidentified militants
- **Action**: Killed
- **Target**: Journalist
- **Location**: Khost city, Khost province
- **Time**: This particular note does not mention anything about time

2.2   Terrorist Attacks

Narratives providing information broadly related to terrorism, but not specifically describing a terrorist attack incident with the characteristics indicated above, should not be considered for this study.

So, let's see examples for a relevant story and a not relevant story for better understanding:

- **Not Relevant:** Although the following story mentions the killing of members of a terrorist organizations, the story is not a relevant one.

  ```
  More than 10 Taliban militants were killed following artillery attack carried
  out by Afghan Security Forces in different parts of Ghorak district on February
  20, reports Bhaktar News.
  ```

  The following story is not describing any terrorist activities. So, the story is not relevant.

  ```
  An Indian Sikh woman was arrested by the Malaysian Police on an unspecified
  date between July to September, 2019 for having links with the Sikhs For
  Justice group, reports The Hindu on September 26.  The arrestee had since
  been deported to India.
  ```

- **Relevant:** The following story talks about a terrorism activity. So, this one is relevant.

  ```
  At least six civilians were killed and six others were injured in an improvised
  explosive device (IED)explosion in an unspecified location in Paktika province
  on January 7, reports Tolo News.
  ```

Following the GTD guidelines, we will only focus on incidents related to terrorist attacks that use any of the following tactics:

- Assassination
- Armed Assault

- Bombing/Explosion

- Hijacking

- Barricade Incident

- Kidnapping

- Facility/Infrastructure Attack

- Unarmed Assault

- Unknown

We recommend you to keep this list handy as you code the news stories.

Below, we replicate the definitions of each type of terrorist attack based on the GTD codebook:

- **Assassination**: An act whose primary objective is to kill one or more specific, prominent individuals. Usually carried out on persons of some note, such as high-ranking military officers, government officials, celebrities, etc. Not to include attacks on non-specific members of a targeted group. The killing of a police officer would be an armed assault unless there is reason to believe the attackers singled out a particularly prominent officer for assassination.

- **Hijacking**: An act whose primary objective is to take control of a vehicle such as an aircraft, boat, bus, etc. for the purpose of diverting it to an unprogrammed destination, force the release of prisoners, or some other political objective. Obtaining payment of a ransom should not the sole purpose of a Hijacking, but can be one element of the incident so long as additional objectives have also been stated. Hijackings are distinct from Hostage Taking because the target is a vehicle, regardless of whether there are people/passengers in the vehicle.

- **Kidnapping**: An act whose primary objective is to take control of hostages for the purpose of achieving a political objective through concessions or through disruption of normal operations. Kidnappings are distinguished from Barricade Incidents (above) in that they involve moving and holding the hostages in another location.

- **Barricade Incident**: An act whose primary objective is to take control of hostages for the purpose of achieving a political objective through concessions or through disruption of normal operations. Such attacks are distinguished from kidnapping since the incident occurs and usually plays out at the target location with little or no intention to hold the hostages for an extended period in a separate clandestine location.

- **Bombing/Explosion**: An attack where the primary effects are caused by an energetically unstablematerial undergoing rapid decomposition and releasing a pressure wave that causes physical damage to the surrounding environment. Can include either high or low explosives (including a dirty bomb) but does not include a nuclear explosive device that releases energy from fission and/or fusion, or an incendiary device where

4

decomposition takes place at a much slower rate. If an attack involves certain classes of explosive devices along with firearms, incendiaries, or sharp objects, then the attack is coded as an armed assault only. The explosive device subcategories that are included in this classification are grenades, projectiles, and unknown or other explosive devices that are thrown in which the bombers are also using firearms or incendiary devices.

- **Armed Assault**: An attack whose primary objective is to cause physical harm or death directly to human beings by use of a firearm, incendiary, or sharp instrument (knife, etc.). Not to include attacks involving the use of fists, rocks, sticks, or other handheld (less-than-lethal) weapons. Also includes attacks involving certain classes of explosive devices in addition to firearms, incendiaries, or sharp instruments. The explosive device subcategories that are included in this classification are grenades, projectiles, and unknown or other explosive devices that are thrown.

- **Unarmed Assault**: An attack whose primary objective is to cause physical harm or death directly to human beings by any means other than explosive, firearm, incendiary, or sharp instrument (knife, etc.). Attacks involving chemical, biological or radiological weapons are considered unarmed assaults.

- **Facility/Infrastructure Attack**: An act, excluding the use of an explosive, whose primary objective is to cause damage to a non-human target, such as a building, monument, train, pipeline, etc. Such attacks include arson and various forms of sabotage (e.g., sabotaging a train track is a facility/infrastructure attack, even if passengers are killed). Facility/infrastructure attacks can include acts which aim to harm an installation, yet also cause harm to people incidentally (e.g. an arson attack primarily aimed at damaging a building, but causes injuries or fatalities).

- **Unknown**: The attack type cannot be determined from the available information.

2.3   Terrorism Targets

To identify the targets of a terrorist attack, we also follow the GTD guidelines related to the types of victims. These incude:

- Business
- Government (general)
- Police
- Military
- Abortion Related
- Airports and Aircraft
- Government (diplomatic)
- Educational Institutions
- Food or Water Supply
- Journalists and Media

- Maritime

- NGO

- Private citizens and property

- Religious figures or institutions

- Telecommunications

- Terrorists / Non-State Militias

- Tourists

- Transportation (other than aviation)

- Unknown

- Utilities

- Violent Political Parties

Below, you will find the definition of what is included (or not) in each target category.

- **Business**: Businesses are defined as individuals or organizations engaged in commercial or mercantile activity as a means of livelihood. Any attack on a business or private citizens patronizing a business such as a restaurant, gas station, music store, bar, café, etc.

  This includes attacks carried out against corporate offices or employees of firms like mining companies, or oil corporations. Furthermore, includes attacks conducted on business people or corporate officers. Included in this value as well are hospitals and chambers of commerce and cooperatives.

  Does not include attacks carried out in public or quasi-public areas such as "business district or commercial area", or generic business-related individuals such as "businessmen" (these attacks are captured under "Private Citizens and Property", see below.) Also does not include attacks against generic business-related individuals such as "businessmen." Unless the victims were targeted because of their specific business affiliation, these attacks belong in "Private Citizens and Property."

- **Government (general)**: Any attack on a government building; government member, former members, including members of political parties in official capacities, their convoys, or events sponsored by political parties; political movements; or a government sponsored institution where the attack is expressly carried out to harm the government.

  This value includes attacks on judges, public attorneys (e.g., prosecutors), courts and court systems, politicians, royalty, head of state, government employees (unless police or military), election-related attacks, or intelligence agencies and spies.

  This value does not include attacks on political candidates for office or members of political parties that do not hold an elected office (these attacks are captured in "Private Citizens and Property").

- **Police**: This value includes attacks on members of the police force or police installations; this includes police boxes, patrols headquarters, academies, cars, checkpoints, etc. Includes attacks against jails or prison facilities, or jail or prison staff or guards.

- **Military**: Includes attacks against military units, patrols, barracks, convoys, jeeps, and aircraft. Also includes attacks on recruiting sites, and soldiers engaged in internal policing functions such as at checkpoints and in anti-narcotics activities. This category also includes peacekeeping units that conduct military operations (e.g., AMISOM)

  Excludes attacks against non-state militias and guerrillas, these types of attacks are coded as "Terrorist/Non-state Militias" see below.

- **Abortion Related**: Attacks on abortion clinics, employees, patrons, or security personnel stationed at clinics.

- **Airports and Aircraft**: An attack that was carried out either against an aircraft or against an airport. Attacks against airline employees while on board are also included in this value.

  Includes attacks conducted against airport business offices and executives. Military aircraft are not included.

- **Government (diplomatic)**: Attacks carried out against foreign missions, including embassies, consulates, etc.

  This value includes cultural centers that have diplomatic functions, and attacks against diplomatic staff and their families (when the relationship is relevant to the motive of the attack) and property. The United Nations is a diplomatic target.

- **Educational Institutions**: Attacks against schools, teachers, or guards protecting school sites. Includes attacks against university professors, teaching staff and school buses. Moreover, includes attacks against religious schools in this value.

  As noted below in the "Private Citizens and Property" value, the GTD has several attacks against students. If attacks involving students are not expressly against a school, university or other educational institution or are carried out in an educational setting, they are coded as private citizens and property.

  Excludes attacks against military schools (attacks on military schools are coded as "Military," see below).

- **Food or Water Supply**: Attacks on food or water supplies or reserves are included in this value. This generally includes attacks aimed at the infrastructure related to food and water for human consumption.

- **Journalists and Media**: Includes, attacks on reporters, news assistants, photographers, publishers, as well as attacks on media headquarters and offices.

  Attacks on transmission facilities such as antennae or transmission towers, or broadcast infrastructure are coded as "Telecommunications," see below.

- **Maritime (includes ports and maritime facilities)**: Includes civilian maritime: attacks against fishing ships, oil tankers, ferries, yachts, etc. (Attacks on fishermen are coded as "Private Citizens and Property," see below).

- **NGO**: Includes attacks on offices and employees of non-governmental organizations (NGOs). NGOs here include large multinational non-governmental organizations such as the Red Cross and Doctors without Borders, as well as domestic organizations.

  Does not include labor unions, social clubs, student groups, and other non-NGO (such cases are coded as "Private Citizens and Property", see below)

- **Other**: This value includes acts of terrorism committed against targets which do not fit into other categories. Some examples include ambulances, firefighters, and international demilitarized zones.

- **Private citizens and property**: This value includes attacks on individuals, the public in general or attacks in public areas including markets, commercial streets, busy intersections and pedestrian malls.

  Also includes ambiguous cases where the target/victim was a named individual, or where the target/victim of an attack could be identified by name, age, occupation, gender or nationality. This value also includes ceremonial events, such as weddings and funerals.

  The GTD contains a number of attacks against students. If these attacks are not expressly against a school, university or other educational institution or are not carried out in an educational setting, these attacks are coded using this value. Also, includes incidents involving political supporters as private citizens and property, provided that these supporters are not part of a government-sponsored event.

  Finally, this value includes police informers. Does not include attacks causing civilian casualties in businesses such as restaurants, cafes or movie theaters (these categories are coded as "Business" see above).

- **Religious figures or institutions**: This value includes attacks on religious leaders, (Imams, priests, bishops, etc.), religious institutions (mosques, churches), religious places or objects (shrines, relics, etc.). This value also includes attacks on organizations that are affiliated with religious entities that are not NGOs, businesses or schools.

  Attacks on religious pilgrims are considered "Private Citizens and Property;" attacks on missionaries are considered religious figures.

- **Telecommunications**: This includes attacks on facilities and infrastructure for the transmission of information. More specifically this value includes things like cell phone towers, telephone booths, television transmitters, radio, and microwave towers.

- **Terrorists / Non-State Militias**: Terrorists or members of identified terrorist groups within the GTD are included in this value. Membership is broadly defined and includes informants for terrorist groups, but excludes former or surrendered terrorists.

  This value also includes cases involving the targeting of militias and guerrillas.

- **Tourists**: This value includes the targeting of tour buses, tourists, or "tours." Tourists are persons who travel primarily for the purposes of leisure or amusement. Government tourist offices are included in this value.

The attack must clearly target tourists, not just an assault on a business or transportation system used by tourists. Travel agencies are coded as business targets.

- **Transportation (other than aviation)**: Attacks on public transportation systems are included in this value. This can include efforts to assault public buses, minibuses, trains, metro/subways, highways (if the highway itself is the target of the attack), bridges, roads, etc. The GTD contains a number of attacks on generic terms such as "cars" or "vehicles."

  These attacks are assumed to be against "Private Citizens and Property" unless shown to be against public transportation systems. In this regard, buses are assumed to be public transportation unless otherwise noted.

- **Unknown**: The target type cannot be determined from the available information.

- **Utilities**: This value pertains to facilities for the transmission or generation of energy. For example, power lines, oil pipelines, electrical transformers, high tension lines, gas and electric substations, are all included in this value. This value also includes lampposts or street lights.

  Attacks on officers, employees or facilities of utility companies excluding the type of facilities above are coded as business.

- **Violent Political Parties**: This value pertains to entities that are both political parties (and thus, coded as "government" in this coding scheme) and terrorists. It is operationally defined as groups that engage in electoral politics and appear as "Perpetrators" in the GTD.

3 Annotation

Quality is not a coincidence!. It is very important to keep the highest possible standards of quality in each step of this research. Part of the reason is that the annotations that coders make at this stage, will be consequential for all the subsequent stages of this project. In particular, we will use human annotations to train a Machine Learning (ML) model. So, if the human annotation is poor quality, then the ML output is going to be unreliable. For this reason, it is important that coders carefully read this manual and familiarize themselves with the concepts and categories. This codebook should serve as a constant point of reference in the coding process. In addition, we will have weekly meetings to clarify questions and discuss specific challenges.

    TagTog. To conduct these annotations, we will rely on `tagtog`, a free web interface that allows for easy and intuitive manual annotation of text. Before starting, please open an account in `https://www.tagtog.net/`. Once you are in, we will send you a link to the collection of SATP stories.

Coder assignment: Please work on the following assigned folders:

- Training stage. Each of the following folders contain 100 stories each:

– Karla: `https://www.tagtog.net/ssp180002/SATP_Practice_1/pool`

– Jessica: `https://www.tagtog.net/ssp180002/SATP_Practice_2/pool`

– Karina: `https://www.tagtog.net/ssp180002/SATP_Practice_3/pool`

- Full coding. Each folder containing 500 stories:

  – Karla:
    * Folder 1 - Folder `TBD`
    * Folder 2 - Folder `TBD`

  – Jessica:
    * Folder 1 - Folder `TBD`
    * Folder 2 - Folder `TBD`

  – Karina:
    * Folder 1 - Folder `TBD`
    * Folder 2 - Folder `TBD`

Each of the coders will get 2 project folders from tagtog. Inside, each of these folders contains 500 stories. Figure 2 is an inside look of a project folder.

Figure 2: tagtog - landing page of a project

Each row represents a single story. If you click on any of them you will be able to view the story like the one shown in Figure 3.

Figure 3: tagtog - a single story view

For this annotation task, we will proceed in three stages. We will work as a group to complete each stage before moving on to the next step. These stages are designed to gradually immerse the coders into the specific content of the stories and the increasing sophistication of the annotation tasks.

1. Relevance

2. Event Coding

3. Action Classification

3.1   Relevance

The first task consists in identifying whether or not a narrative is relevant for the purposes of our study. A narrative is considered relevant if it provides specific information about a terrorists attack including the tactics described in section 2.2. As a reminder, these actions include:

- Assassination
- Armed Assault
- Bombing/Explosion
- Hijacking
- Barricade Incident

- Kidnapping

- Facility/Infrastructure Attack

- Unarmed Assault

- Unknown

Inside tagtog, I have created some entities and a label. Also, for the Action events has its own set of labeling!

- **Is Relevant** : This is a document level label. Here, if the document/story is related to terrorist activity then you have to select:

  - True

  - False

To classify the stories, you need to perform the following steps:

1. Go inside a story (see Figure 4), you will see a drop down box under the Document Labels section.

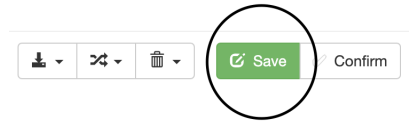Figure 4: Position of the "Is Relevant" field
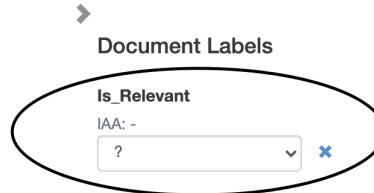
2. Click on the drop down box, you will be able to see two values (see Figure 5) . So, if the story is relevant for the terrorist event you will select true, or false otherwise.

Figure 5: Values of the "Is Relevant" field of a story
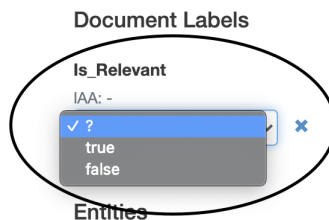
3. Click on Save (see Figure 6).

Figure 6: Save the "Is Relevant" field of a story

3.2   Event Coding

As discussed above, an event is characterized by having the following elements: source, action, target, location, date.

As such, the second task consists of labeling the main components of an event in each of the news stories. To do so, we configured the following labels in tagtog:

- **Action-1, Action-2, Action-3**: A story can have multiple events. You will consider at most 3 events. So, each of the event will talk about separate actions. So, you have to use these entities to annotate the separate actions of the events.

- **Target-1, Target-2, Target-3**: As there could at most 3 events- target set can also be up to 3. So, you will use these target entities to label with its own event.

- **Source-1 , Source-2 , Source-3**: There can be 3 different source as well if there are 3 events in a story. These are the created entities for Source inside tagtog.

- **Location-1 , Location-2 , Location-3**: There can be 3 different locations as well if there are 3 events in a story. These are the created entities for Location inside tagtog.

- **Date-1 , Date-2 , Date-3**: There can be 3 different dates as well if there are 3 events in a story. These are the created entities for dates inside tagtog.

To code the events (Action, Target, Source, Location, Date) you have to take the following steps. Also, in a story we can find more than one event. So, in that case we will take an account of maximum 3 events. All these events will have its own action, target, source, location and date. You have to annotate them respectively.

1. You have to select the text that you want to annotate as in Figure Figure 7. Let's select the keyword "civilians" as it is a target.

Figure 7: tagtog - select the text for the event component

2. Once you have selected your text, if you go to the selected part and left click on it, as Figure Figure 8, you will be able to select the event type respectively. In the example, "civilians" are "Target-1". So, we will select "Target-1".

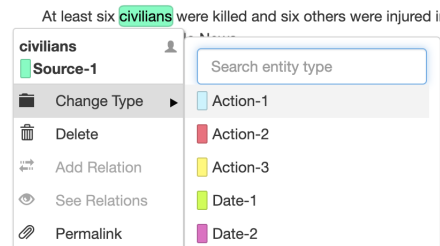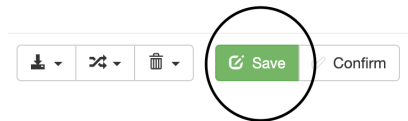Figure 8: tagtog - annotate the text to correct event

Figure 9: tagtog - annotated Target-1

3. Once you have selected "Target-1", it will get assigned as one of the value for "Target-1" (Figure 9).

4. Finally, we save the annotation as Figure 10 shows.

Figure 10: tagtog - save the document

3.3 Action Classification

The third classification task refers to annotating the specific type of terrorist attack mentioned in the SATP stories. Please refer back to section 2.2 for the detailed description of GTD terrorist attacks. As a reminder, here is a list of the types of terrorist activities:

- Assassination
- Armed Assault
- Bombing/Explosion
- Hijacking
- Barricade Incident
- Kidnapping
- Facility/Infrastructure Attack
- Unarmed Assault
- Unknown

To classify the action according to the GTD, you have to take the following steps.
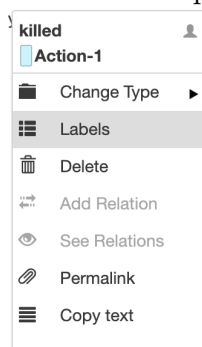
1. Select all the keywords for an Action and code them Action. In this example, the keywords are - "killed", "Improvised Explosive Device(IED Explosion)" and "explosion" and all of them are coded as "Action-1" as these keywords relate to a single action (Figure 11).

Figure 11: tagtog - annotate the actions

At least six civilians were killed and six others were injured in an improvised explosive device (IED) explosion in an unspecified location in Paktika province on January 7, reports Tolo News.
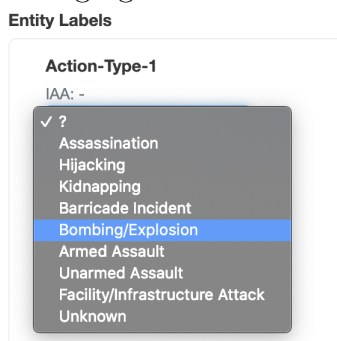
2. After that, select each keyword and perform a left click. Then, you will see an option in the drop down menu called "Labels". Click on that (Figure 12).

Figure 12: tagtog - select labels option from drop down

killed
Action-1

📁 Change Type ▶
☰ Labels
🗑 Delete
⇄ Add Relation
👁 See Relations
📎 Permalink
☰ Copy text

3. Finally, a pop up window will come which will have a drop down box like the following where you can select the type of the action. In this case, the type of the action is "Bombing/Explosion" (Figure 13). So, you need to select that option.

Figure 13: tagtog - select label for actions

Entity Labels

Action-Type-1
IAA: -
✓ ?
Assassination
Hijacking
Kidnapping
Barricade Incident
Bombing/Explosion
Armed Assault
Unarmed Assault
Facility/Infrastructure Attack
Unknown

4. Last but not the least - whenever you modify or add any annotation, you need to save it. The good thing about tagtog is, when you modify the story the "Save" button (on the upper right) turns into green. Once you are done annotating, don't forget to hit on the "save" button (Figure 14).

Figure 14: tagtog - save the document
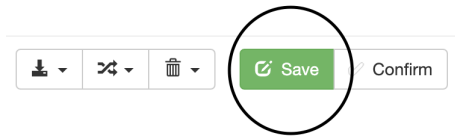
4 Meetings and communications

Please mark your calendars for our weekly meeting on Mondays at 9:00 am via zoom. It is crucial for you not to miss any meeting as this is our main opportunity to provide directions and clarify questions.

Also, please use Slack for regular communications. It is better to use the `general` channel for work-related questions so everyone is aware of the conversations we are having. You can also use direct messages if you need some private communications.