World news classification

How to classify world news

All the participants are required to read the following coding manual and fully understand the coding rules. The coding rules are simple and clear, but please do not participate in this study if you find the following rules too complex to understand.

Introduction

In this study, all codes for news classification are provided in drop-down lists in the online platform, and will consist of country, region and "I don't know" codes.

The principle of news classification is focus on locations of events and issues in stories. For example, the following story has to be coded as Japan because this story is about an even took place in Japan:

> **Japan plant explosion kills at least five**
>
> AN explosion at a chemical factory in Japan has killed at least five people and injured a dozen more, police have revealed.

In this way, we are often able to determine which countries news stories are associated with.

However, there are more complex stories that require elaborated rules. So the classification rules of this study explained in this instruction contain both **principle** and **exceptions**. Participants first try to classify a story based on the **principle**, and follow **exceptions** only if the principle does not help to identify a country that is most strongly associated with story.

All the participants are required to read through this instruction carefully before starting their work, but you can always return to this instruction during the course of the work to perform accurate classification.

Principle: Focus on events and issues

The first example was easy to classify because the story mentions only one country (Japan), but more than one country are often mentioned in international news stories. In those cases, we focus on locations of events or issues that the story is concerned. This means that news stories should not be classified based on actors' association with countries. For example, the next news story has to be coded as Tunisia, but not as the United States.

> **Kerry Makes Surprise Stop in Tunisia**
>
> The visit by Secretary of State John Kerry was aimed at demonstrating American support for Tunisia's transition to democracy.

Similarly, the following story has to be coded as Libya, because the kidnapping had happened in the

Libyan city of Tripoli:

> **Egyptian Envoys Plead for Release**
>
> Five Egyptian diplomats kidnapped in Tripoli in retaliation for Egypt's arrest of a Libyan militia chief pleaded on Saturday for their government to free him to secure their release.

If a news story is reporting reactions of public officials to a foreign event, it has to be coded according to the location of the original event. For instance, the next news story has to be coded as Ukraine, despite the location of the meeting is Hague in Netherland, because it is a response to the Ukraine crisis:

> **World leaders meet as Ukraine crisis strains East-West ties**
>
> World leaders gather in The Hague on Monday for talks set to be dominated by the crisis in Ukraine, with Russia facing possible exclusion from the G8 club of rich nations as punishment for its actions.

The next story has to be coded as Germany, because the original even is Germany's planning, and the Romanian Prime Minister is only responding to the event:

> **German plan to fingerprint Roma draws 'Nazi' gibe**
>
> Romania's Prime Minister has accused a top German politician of having a "Nazi mindset" for suggesting that his countrymen should...

If news story is reporting a course of event, and the event is associated with different locations, we have to choose the most important country. The next story has to be coded as Somalia because the hostages have been in Somalia for four years and they are only temporarily staying in Kenya:

> **Sailors held by pirates escape after four years**
>
> Eleven sailors mostly from Asia held hostage for almost four years by Somali pirates escaped their captors and are safe in Kenya, mediators who helped secure their freedom said Saturday.

Next story has to be coded as Lebanon, because the refugees are currently in Lebanon and Syria is only their origin:

> **Czech Republic ready to support Syrian refugees in Lebanon**
>
> Czech Republic foreign minister Lubomir Zaoralek said on Wednesday that his country is willing to help Lebanon to offer assistance to the Syrian refugees at are taking shelters in the country.

Some news stories do not contain names of countries or cities, but they mention institutions, groups or public figures that are association with certain locations. If participants cannot relate those actors to particular countries, a quick online search has to be performed (e.g. Wikipedia). For example, the next story does not refer to any country, but participants can easily find that Nusra Front operates in Syria and it has to be coded as Syria accordingly:

> **Nusra Front threaten vengence for air strikes**
>
> Militants in the al-Qaeda linked Nusra Front have said that coalition air strikes are a "war against Islam" as it...

For the next story, by using the Internet, participants can easily find that Rubin Carter had both American and Canadian citizenship but (wrongly) convicted in the United States. Hence, this story has to be coded as the United States:

> **'Hero' Hurricane fought for justice**
>
> RUBIN "Hurricane" Carter, the prize-fighter whose wrongful murder conviction made him a symbol of injustice, has died.

Exception 1: Focus on actors

If there is no particular event or issue in stories, and statements of a public figures are the main content, they have to be coded based on the actors' association. The next story has to be coded as the United States, because there is no mention to event or issue at all, and the statement was made by the President of the United State:

> **UK and France equally special, says Obama**
>
> Britain received equal billing with France from President Obama yesterday as Washington celebrated a renaissance in its friendship with Paris....

In the same way, the following story also has to be coded as Bangladesh, since it is not referring to any particular event in China, and the main content is the comment made by the Bangladesh's Prime Minister on the country's independence:

> **World politics influenced China's stance in 1971: Hasina**
>
> Prime Minister Sheikh Hasina on Saturday sought to downplay the Chinese opposition to Bangladesh's independence in 1971, saying world politics had influenced China's stance at that time.

Actors to be focused are not limited to public officials, but other group of people. For instance, the following story has to be coded as Sri Lanka, because the story is about accusation against the Pope

by a Buddhist group in Sri Lanka.

> **Hardline Buddhists want Pope Francis to apologize**
>
> A Buddhist group accused of instigating recent attacks on Muslims in Sri Lanka says Pope Francis must apologize to Buddhists for atrocities allegedly committed by Christian colonial rulers of the South Asian island nation when he visits next year.

Exception 2: Code as regions

Sometimes, due to the international nature of the evens or issues, participants are not able to establish a link between a story and a particular country. For those cases, there are 7 regional codes that can be used: Africa, America, Antarctic, Arctic, Asia, Europe, Oceania. For example, the stories presented below can be coded respectively as Europe and Antarctic:

> **European Border Agency Reports Surge in Illegal Migration**
>
> The report by the European Union's border agency feeds a debate on immigration that has led to political gains by far-right parties in the European Union.

> **US drawn into pack ice saga**
>
> THE Antarctic stranding of an Australian-sponsored "pseudo-scientific" tourism venture has become a sensitive diplomatic issue.

Exception 3: No information about locations

Not all news stories contain names of places or actors. Some news stories are strongly related to certain places, but do not contain information about the locations just like the next example. Participants have to code those stories as "I don't know, because there is no information about location":

> **Luxury shantytown where rich can pretend to be poor**
>
> For most tourists, a shantytown complete with open, fly-infested outside toilets would hardly be the ideal choice for a romantic...

Exception 4: No association with locations

By the nature of events or issues, it is sometimes impossible to find association with a particular country or region, even when there are mentions of actors or places in news stories. The following three example are this type of stories and have to be coded as "I don't know, because there is no association with location":

> **Light can escape from black hole: Hawking**
>
> STEPHEN Hawking has suggested that black holes are not always the final graveyard for the light and matter sucked into them.

> **Scientists pinpoint 'gay gene'**
>
> SCIENTISTS have found the strongest evidence yet for the existence of genes that increase the chance of a man being gay.

> **World Briefing: U.N. Council Resolution Urges No Ransom to Terrorists**
>
> The Security Council unanimously passed a resolution pressing governments around the world not to pay ransom money to terrorist kidnappers.

Exception 5: Technical problem in text

Finally, this category is rarely used, but still possible. Due to the deficiency in the data used in this study, participants might find news story that do not make sense or are even unreadable. In such cases, participants have to report the problem by coding as "I don't know, because there is a technical problem in text". For instance, the next news story does not make sense because there is a mismatch between the heading and the lead:

> **Afghanistan parliament approves US, Nato agreements**
>
> NRI industrialist Lord Swraj Paul has said that he "feels very sad" that the British government cannot find little money for the proposed Mahatma Gandhi statue to be installed at the Parliament Square here.