

Computational Approaches in the Study of Human Rights: Measuring Human Rights Conditions

Baekkwan Park

East Carolina University

Machine Learning and Social Sciences Workshop:
Sungkyunkwan University

Feb 19, 2021

Human Rights Organizations (HROs): **Reviewers**



HUMAN
RIGHTS
WATCH



ICTJ | Justice
Truth
Dignity

apt

la prévention de la
association for the prevention
asociación para la prevención



icj

International
Commission
of Jurists



fidh
International Federation
for Human Rights



ISHR

Human Rights Organizations (HROs)



29 OCTOBER 2018

More than 150 human rights defenders gather to lead change and fight repression, racism and discrimination



ICRC



30 OCTOBER 2018

Maldives: The tortuous ordeal of a prisoner in paradise



29 OCTOBER 2018

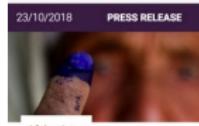
Eswatini: New Prime Minister must ensure respect for human rights

29 OCTOBER 2018

SAUDI ARABIA

UPDATED: 29 OCTOBER 2018

Leading journalists demand justice for Jamal Khashoggi



Afghanistan

Voters deserve praise but serious election flaws must be urgently addressed



Human Rights Watch



Human Rights Defenders
Ukraine: Adoption of "Foreign Agents" threaten human

fidh



Human Rights Defenders
Singapore: Jolovan Wham convicted for "scandalising the court"



Human Rights Defenders
Turkey: lift of travel ban against Eren Keskin



October 16, 2018 | Witness
Witness: Girls in Senegal Risk Losing Their Futures



Be Like Nature

Surgeries on Intersex

"It's Like We're Always in a Prison"

Abuses Against Boys Accused of National Security Offenses in Somalia



A prison warden at a prison in Garowe, Puntland state,

"Like Living in Hell"
Police Abuses Against Child and Adult Migrants in Calais

Available in English Français



17 years later, Guantánamo is still a threat to human rights
The military prison remains a stain on the human rights record of the United States
amnesty.org

HROs publish “reviews (evaluation)”: *press releases, campaign reports, policy reports, facebook, twitter, etc.*

Human Rights Scores



Composing



State
Department

Coding

Social
Scientists

3

PTS



Reports to Scores (PTS)



THE POLITICAL TERROR SCALE

ABOUT NEWS DATA BIBLIOGRAPHY



The Political Terror Scale

[Download the Data](#)

About the PTS

Learn about the Political Terror Scale project, the team of researchers, and how we measure and code human rights violations.

News & Updates

Find the latest news and updates for The Political Terror Scale data collection project — as well as links to the latest publications by members of the PTS research team.

Data and Download

Explore and download the latest release of the PTS and SVS data — covering human rights rights records from 1976 to 2014.

Bibliography

Browse our bibliography of scholarly publications studying human rights by means of the Political Terror Scale data.



Reports to Scores (PTS)

Political Terror Scale Levels

Level	Interpretation
1	Countries under a secure rule of law, people are not imprisoned for their views, and torture is rare or exceptional. Political murders are extremely rare.
2	There is a limited amount of imprisonment for nonviolent political activity. However, few persons are affected, torture and beatings are exceptional. Political murder is rare.
3	There is extensive political imprisonment, or a recent history of such imprisonment. Execution or other political murders and brutality may be common. Unlimited detention, with or without a trial, for political views is accepted.
4	Civil and political rights violations have expanded to large numbers of the population. Murders, disappearances, and torture are a common part of life. In spite of its generality, on this level terror affects those who interest themselves in politics or ideas.
5	Terror has expanded to the whole population. The leaders of these societies place no limits on the means or thoroughness with which they pursue personal or ideological goals.

Reports to Human Rights Scores

The screenshot shows the official website of the U.S. Department of State's Bureau of Democracy, Human Rights, and Labor. The header features the seal of the United States and the text "U.S. DEPARTMENT OF STATE DIPLOMACY IN ACTION". A search bar at the top right contains the placeholder "Search Site...". Below the header, a sidebar titled "In This Section:" lists years from 1999 to 2017. The main content area includes a breadcrumb trail ("Home > Under Secretary for Civilian Security, Democracy, and Human Rights > Bureau of Democracy, Human Rights, and Labor > Releases"), a "Human Rights Reports" section with a detailed description of the annual reports, a "Current Reports" section, and a link to the "Bureau of Democracy, Human Rights, and Labor homepage". Social media sharing icons are present at the top right of the main content.

US State Department

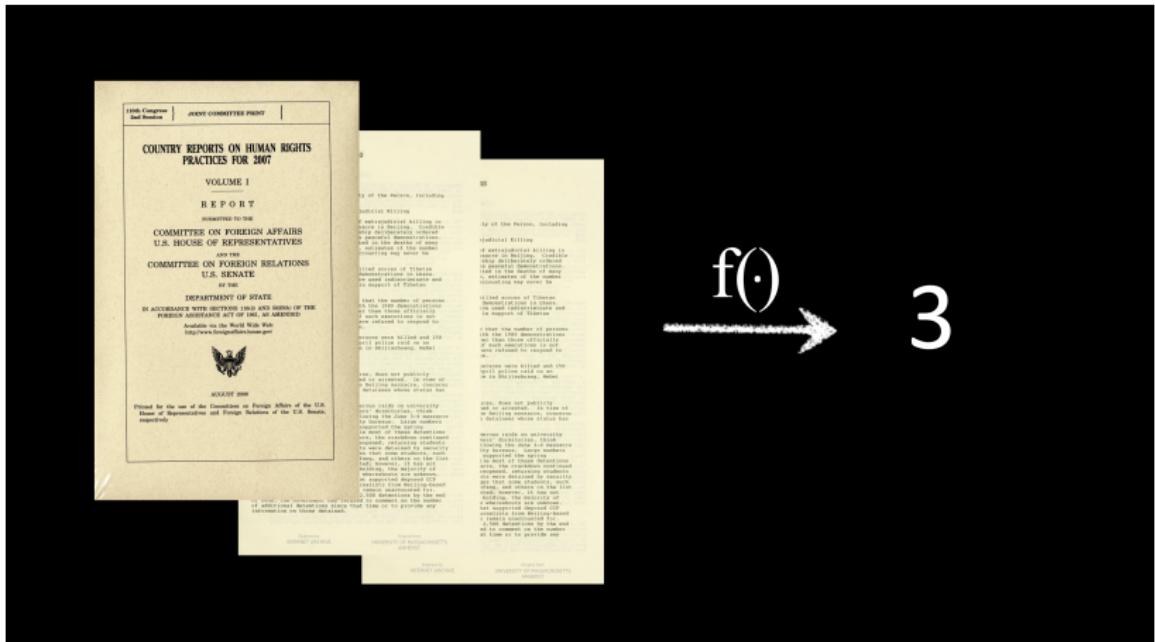
The screenshot shows the homepage of Amnesty International. The header features the organization's logo and navigation links for "TAKE ACTION" and "DONATE". Below the header is a large photograph of a man in a striped shirt sitting in a debris-strewn urban environment. To the left of the photo is a white rectangular overlay containing the word "REPORTS" in bold capital letters. The overall design is clean with a focus on the visual impact of the photograph.

"How-to" resources to help you protect human rights

We produce reports based on rigorous and independent research. These reports document

Amnesty International

Reports to Scores



Reports to Scores: PTS Codebook

Scope, Intensity, and Range

Scope pertains to the type of physical integrity rights violation that is described in a given report (e.g., arbitrary arrest, rape, killings). Violations must be perpetrated by an agent of the state (or at his or her behest) *and* within the territorial boundaries of the state or entity in question. Coders are asked to count an instance of a more severe type of violation (e.g., an extra-judicial killing) more heavily than a less severe one (e.g., an instance of excessive use of force, an arbitrary arrest). *Intensity* refers to the frequency with which violations occur (e.g., isolated and rare instances, common or routine practice). Coders are asked to weigh evidence of multiple violations (e.g., mass arrests) more heavily than a single instance. Similarly, rare or isolated events are to be considered less, compared with reports of systematic abuse. Finally, *range* is the proportion of the population subjected to abuse (e.g., targeted and selective violence, indiscriminate abuse). Coders are expected to assign worse or higher scores if violence is indiscriminate (e.g., affecting the whole country, the entire population) rather than selective (e.g., targeting select members of a particular group).

Coding Procedure

As outlined above, coders are asked to assign a score (1 through 5) to a report based on evidence contained in the reports as to the *scope*, *intensity*, and *range* of physical integrity rights violations perpetrated by state agents within the boundaries of the country (or entity) covered by the report. This is a complex undertaking and requires careful reading and evaluation of reports that often differ vastly in length and detail from country to country, year to year, and importantly from one reporting agency (i.e., Amnesty International, Human Rights Watch, and the US Department of State) to another.

Identifying Allegations

For every allegation contained in a report coders must first determine whether the allegation falls within the **Scope** of the **PTS**. This task typically requires coder to answer the following **four** questions:

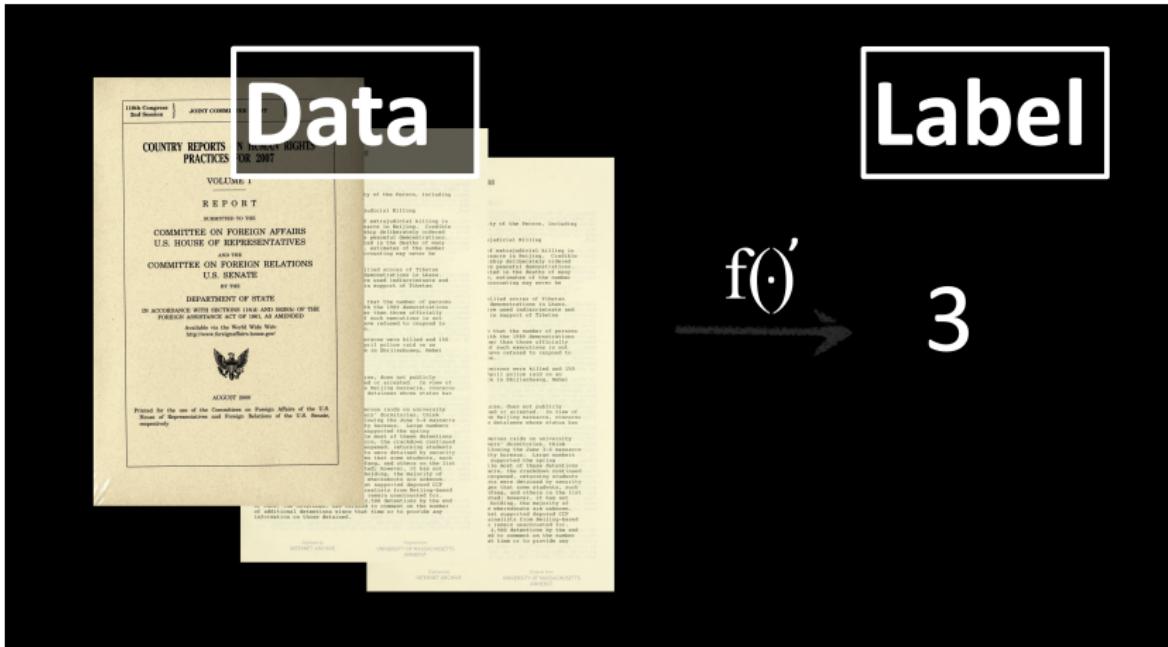
1. Violation Timing Requirement

Was the alleged act or violation perpetrated during the year covered by the report? See:
Unit of Observation: Timing

Example: US State Department Report, Turkey 2019



Reports to Scores



Converting Texts to Numbers

- S_1 : *The government committed unlawful killings.*
- S_2 : *The government agents committed arbitrary killings.*

$$\mathbf{S} = \begin{bmatrix} \text{agents} & \text{arbitrary} & \text{committed} & \text{government} & \text{killings} & \text{the} & \text{unlawful} \\ 0 & 0 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 & 1 & 0 \end{bmatrix} \begin{matrix} S_1 \\ S_2 \end{matrix}$$

Mapping Function?

$$\mathbf{D} = \begin{bmatrix} w_1 & w_2 & w_3 & \dots & w_N \\ 0 & 0 & 1 & \dots & 1 \\ 1 & 1 & 1 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & 1 \end{bmatrix} \begin{matrix} D_1 \\ D_2 \\ \vdots \\ D_K \end{matrix}$$
$$\mathbf{Y} = \begin{bmatrix} y_n \\ 0 \\ 1 \\ \vdots \\ 0 \end{bmatrix} \begin{matrix} y_1 \\ y_2 \\ \vdots \\ y_K \end{matrix}$$

Estimating Mapping Function

$$z = \left(\sum_{i=1}^n w_i x_i \right) + b$$

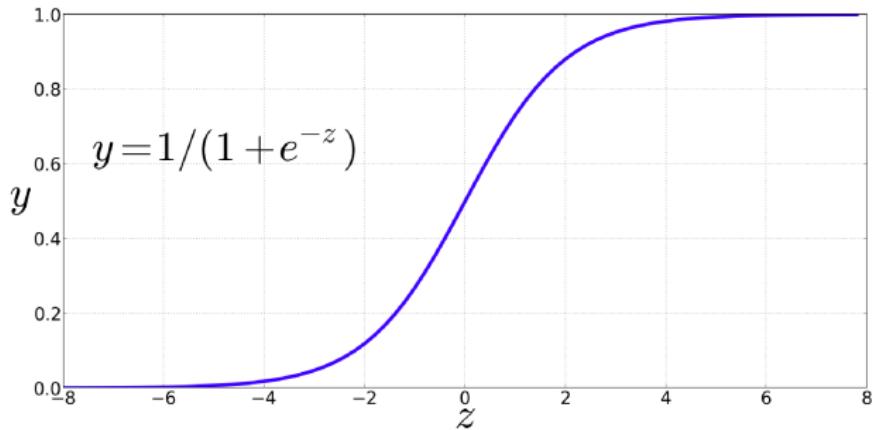
$$z = w \cdot x + b$$

Estimating Mapping Function

$$P(y = 1|x)$$

$$P(y = 0|x)$$

Estimating Mapping Function



Estimating Mapping Function

$$y = \sigma(z) = \frac{1}{1 + e^{-z}} = \frac{1}{1 + \exp(-z)}$$

Estimating Mapping Function

$$\begin{aligned}P(y = 1) &= \sigma(w \cdot x + b) \\&= \frac{1}{1 + e^{-(w \cdot x + b)}}\end{aligned}$$

$$\begin{aligned}P(y = 0) &= 1 - \sigma(w \cdot x + b) \\&= 1 - \frac{1}{1 + e^{-(w \cdot x + b)}} \\&= \frac{e^{-(w \cdot x + b)}}{1 + e^{-(w \cdot x + b)}}\end{aligned}$$

Loss Function

$L(\hat{y}, y)$ = How much \hat{y} differs from the true y

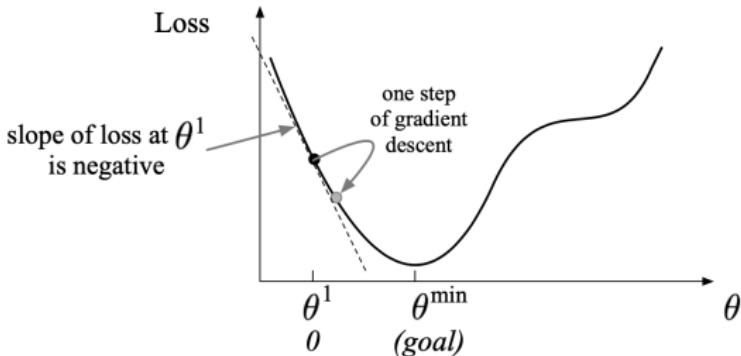
$$\hat{\theta} = \arg \min_{\theta} \frac{1}{m} \sum_{i=1}^m L(f(x^{(i)}; \theta), y^{(i)})$$

$$\nabla_{\theta} L(f(x; \theta), y) = \begin{bmatrix} \frac{\partial}{\partial w_1} L(f(x; \theta), y) \\ \frac{\partial}{\partial w_2} L(f(x; \theta), y) \\ \vdots \\ \frac{\partial}{\partial w_n} L(f(x; \theta), y) \end{bmatrix}$$

$$\theta^{t+1} = \theta^t - \eta \nabla_{\theta} L(f(x^{(i)}; \theta), y^{(i)})$$

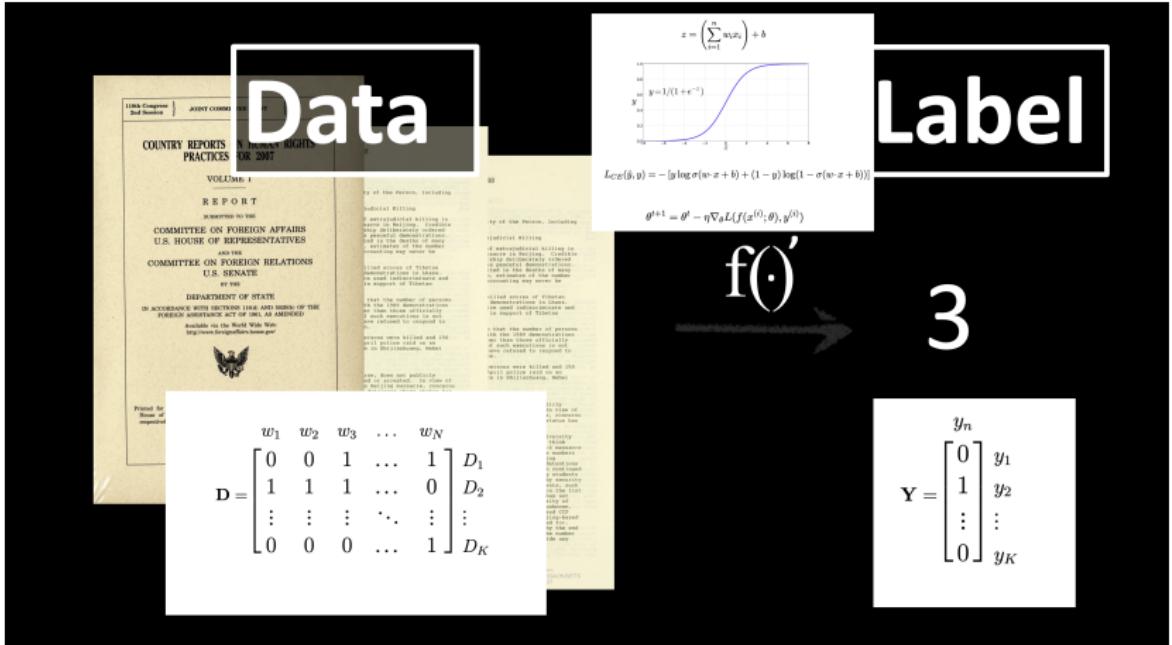
$$L(\hat{y}, y) = -[y \log \sigma(w \cdot x + b) + (1 - y) \log(1 - \sigma(w \cdot x + b))]$$

Optimization: Gradient Descent

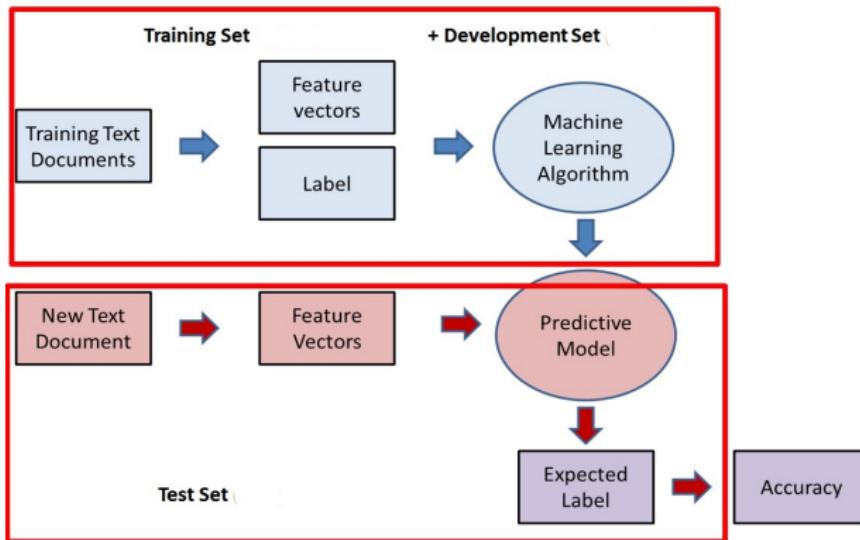


$$\theta^{t+1} = \theta^t - \eta \nabla_{\theta} L(f(x^{(i)}; \theta), y^{(i)})$$

Reports to Scores



ML Training and Testing: Data Split



Easy Way

learn Install User Guide API Examples More ▾

scikit-learn

Machine Learning in Python

Getting Started Release Highlights for 0.24 GitHub

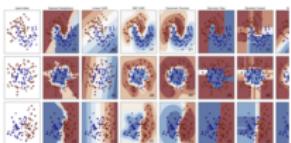
- Simple and efficient tools for predictive data analysis
- Accessible to everybody, and reusable in various contexts
- Built on NumPy, SciPy, and matplotlib
- Open source, commercially usable - BSD license

Classification

Identifying which category an object belongs to.

Applications: Spam detection, image recognition.

Algorithms: SVM, nearest neighbors, random forest, and more...



Examples

Dimensionality reduction

Reducing the number of random variables to consider.

Applications: Visualization, Increased efficiency

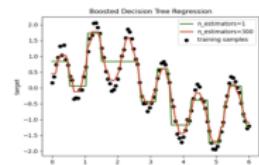
Algorithms: k-Means, feature selection, non-negative matrix factorization, and more...

Regression

Predicting a continuous-valued attribute associated with an object.

Applications: Drug response, Stock prices.

Algorithms: SVR, nearest neighbors, random forest, and more...



Examples

Clustering

Automatic grouping of similar objects into sets.

Applications: Customer segmentation, Grouping experiment outcomes

Algorithms: k-Means, spectral clustering, mean-shift, and more...



Examples

Model selection

Comparing, validating and choosing parameters and models.

Applications: Improved accuracy via parameter tuning

Algorithms: grid search, cross validation, metrics,

Preprocessing

Feature extraction and normalization.

Applications: Transforming input data such as text for use with machine learning algorithms.

Algorithms: preprocessing, feature extraction, and more...