

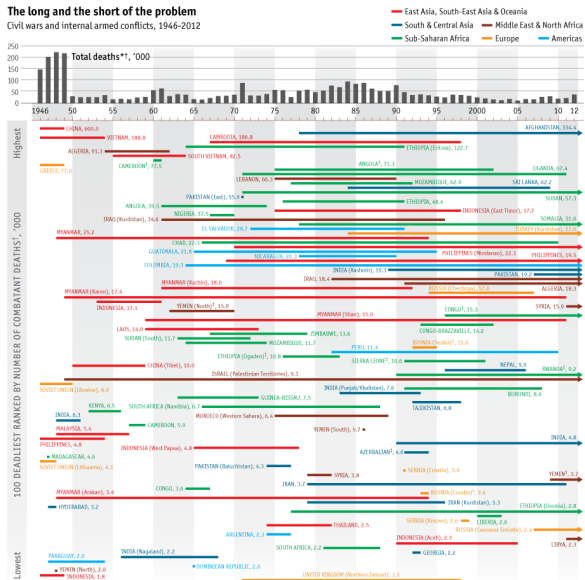
Preread slides for Thursday, Sept 24:
Armed conflict, part 2

Matthew J. Salganik

COS 597E/SOC 555 Limits to prediction
Fall 2020, Princeton University

The long and the short of the problem

Civil wars and internal armed conflicts, 1946-2012



Sources: PRIO; Uppsala University

*Based on over 250 conflicts, 1946-2012. †Deaths in battle of government troops and troops of politically organised rebels; conflicts restarted within 10 years counted as continuous. ‡Including foreign intervention

Comparing Random Forest with Logistic Regression for Predicting Class-Imbalanced Civil War Onset Data

David Muchlinski

School of Social and Political Science, University of Glasgow, Glasgow, UK
e-mail: david.muchlinski@glasgow.ac.uk (corresponding author)

David Siroky

Department of Political Science, Arizona State University, Tempe, AZ
e-mail: david.siroky@asu.edu

Jingrui He

Department of Computer Science and Engineering, Arizona State University, Tempe, AZ
e-mail: jingrui.he@asu.edu

Matthew Kocher

Department of Political Science, Yale University, New Haven, CT
e-mail: mathew.kocher@yale.edu

Reading notes:

- ▶ Major goals: (1) compare random forest to logistic regression for predicting civil war onset (2) learn from random forest about civil war onset.

Reading notes:

- ▶ Major goals: (1) compare random forest to logistic regression for predicting civil war onset (2) learn from random forest about civil war onset.
- ▶ For goal (1), note that two things are varying: number of predictors and learning algorithm.

Reading notes:

- ▶ Major goals: (1) compare random forest to logistic regression for predicting civil war onset (2) learn from random forest about civil war onset.
- ▶ For goal (1), note that two things are varying: number of predictors and learning algorithm.
- ▶ For goal (1), think about the role of time in assessing predictions.

Reading notes:

- ▶ Major goals: (1) compare random forest to logistic regression for predicting civil war onset (2) learn from random forest about civil war onset.
- ▶ For goal (1), note that two things are varying: number of predictors and learning algorithm.
- ▶ For goal (1), think about the role of time in assessing predictions.
- ▶ For goal (2), ask yourself if you believe the results in Fig 4 and 5. It is OK if you don't have strong feelings either way.

Reading notes:

- ▶ Major goals: (1) compare random forest to logistic regression for predicting civil war onset (2) learn from random forest about civil war onset.
- ▶ For goal (1), note that two things are varying: number of predictors and learning algorithm.
- ▶ For goal (1), think about the role of time in assessing predictions.
- ▶ For goal (2), ask yourself if you believe the results in Fig 4 and 5. It is OK if you don't have strong feelings either way.
- ▶ Note that this starts off like a paper about civil wars and seems to end up like a paper about random forest.

Reading notes:

- ▶ Major goals: (1) compare random forest to logistic regression for predicting civil war onset (2) learn from random forest about civil war onset.
- ▶ For goal (1), note that two things are varying: number of predictors and learning algorithm.
- ▶ For goal (1), think about the role of time in assessing predictions.
- ▶ For goal (2), ask yourself if you believe the results in Fig 4 and 5. It is OK if you don't have strong feelings either way.
- ▶ Note that this starts off like a paper about civil wars and seems to end up like a paper about random forest.
- ▶ Do you see any connection to idea of theoretical limits?

Many responses. Reading notes:

- ▶ Neenhoeffer and Sternberg: How to measure generalization? Are we cheating with our cross-validation? How many training loops are there? Questions about cross-validation go beyond the ones raised here. This might feel like the weeds, but it is hard to quantitatively study the limits of predictability if we cannot clearly and accurately measure predictability.

Many responses. Reading notes:

- ▶ Neenhoeffer and Sternberg: How to measure generalization? Are we cheating with our cross-validation? How many training loops are there? Questions about cross-validation go beyond the ones raised here. This might feel like the weeds, but it is hard to quantitatively study the limits of predictability if we cannot clearly and accurately measure predictability.
- ▶ Don't worry so much about the other responses.

Many responses. Reading notes:

- ▶ Neenhoeffer and Sternberg: How to measure generalization? Are we cheating with our cross-validation? How many training loops are there? Questions about cross-validation go beyond the ones raised here. This might feel like the weeds, but it is hard to quantitatively study the limits of predictability if we cannot clearly and accurately measure predictability.
- ▶ Don't worry so much about the other responses.
- ▶ Note the value of open and reproducible research.

Many responses. Reading notes:

- ▶ Neenhoeffer and Sternberg: How to measure generalization? Are we cheating with our cross-validation? How many training loops are there? Questions about cross-validation go beyond the ones raised here. This might feel like the weeds, but it is hard to quantitatively study the limits of predictability if we cannot clearly and accurately measure predictability.
- ▶ Don't worry so much about the other responses.
- ▶ Note the value of open and reproducible research.
- ▶ Do you find the authors' reply convincing?

Preread slides for Thursday, Sept 24:
Armed conflict, part 2

Matthew J. Salganik

COS 597E/SOC 555 Limits to prediction
Fall 2020, Princeton University