Probabilistic Conflict Forecasting with Automated Machine Learning*

Vito D'Orazio

Associate Professor of Political Science and Data Sciences

West Virginia University

Abstract

This contribution is at the grid level. The idea is to generate point predictions using automated machine learning, then do a grid-search over parameterizations of different distributions to find the one that performs best. Models were compared using the CRPS across test windows for (1) different Box-Cox transformations on the dependent variable, (2) different sets of predictor variables, and (3) different distributions (Poisson, Negative Binomial, and Tweedie). To make bolder predictions and to make use of different strengths across each setup, ensembles of probabilistic models were built and compared across the different Box-Cox transformations. Two different models were submitted. The *dorazio_log* model forecasts the log of the dependent variable and then back-transforms to the original scale. This model is more conservative but scores a lower CRPS. Other Box-Cox transformations produced higher forecasted values which scored better for some grid-months. The *dorazio_ensemble* model uses forecasts from five different Box-Cox transformations, selecting which to use based on a separate grid-month forecasting model that was trained to predict when to use which transformation. All other factors between these models are the same, including common input features and the Tweedie distribution.

*This paper documents a contribution to the VIEWS Prediction Challenge 2023/2024. Financial support for the Prediction Challenge was provided by the German Ministry for Foreign Affairs. For more information on the Prediction Challenge please see Hegre et al. (Forthcoming) and https://viewsforecasting.org/research/prediction-challenge-2023. Any questions about this paper may be directed to Vito D'Orazio at vito.dorazio@mail.wvu.edu.

The VIEWS problem is to forecast levels of state-based violence for each of the twelve months of a calendar year using data up to and including October from the previous year (Hegre et al. 2021, 2022). The forecasts may be at the country or grid level of aggregation using prio-grid Tollefsen et al. (2012). This contribution is at the prio-grid month level (PGM).

The idea is to generate point predictions using automated machine learning (autoML), then do a grid-search over parameterizations of different distributions to find the one that performs best. Models were compared using the continuous rank probability score across test windows for (1) different Box-Cox transformations on the dependent variable, (2) different sets of predictor variables, and (3) different distributions for the forecasts (Poisson, Negative Binomial, and Tweedie).

Two different models were submitted: *dorazio_log* and *dorazio_ensemble*. For both, point predictions were obtained using the H2O autoML system and a common set of 74 conflict history variables. Each of these predictor variables was constructed from the raw GED event data using only state-based conflict events. Forecast distributions were obtained by identifying grid-specific parameterizations of the Tweedie distribution and taking 1,000 draws from that distribution with a mean equal to the point prediction for the grid. Each model had its own grid-specific parameterization that was optimized by minimizing the continuous rank probability score using the last five years worth of predictions.

The *dorazio_log* model differs from the *dorazio_ensemble* model in one important way. The *dorazio_log* model forecasts the log of the dependent variable (Box-Cox transformation where $\lambda = 0$) with autoML, and then back-transforms to obtain a count. The *dorazio_ensemble* model forecasts five different Box-Cox transformations, obtaining a full set of forecast distributions for all five. Then, one of those five is selected for each grid-month based on a separate forecasting model that was trained to predict when to use which transformation.

The *dorazio_log* model scores a lower mean CRPS (0.442) but produces relatively low forecasted values. The *dorazio_ensemble* model scores a worse CRPS (0.452) but makes bolder forecasts.

1 Point Predictions

Producing the point predictions generally follows the approach taken in D'Orazio and Lin (2022), which was the best performing model in the first VIEWS forecasting competition.

Expanding the Input Features

The first step to generate the point predictions was to expand the initial data provided by VIEWS with additional conflict history variables. To do this, GED version 24.1 was combined with the UCDP Candidate Events Dataset version 24.0.5 (Sundberg and Melander 2013; Hegre et al. 2020). This provided raw GED event data through April 2024.

This data was aggregated to create a number of new variables, primarily those that focused on counts of events. For example, instead of aggregating the number of state-based fatalities in a grid-month, you can aggregate the count of of state-based conflict events. Results have shown that count-based variables improve performance when forecasting future levels of fatalities (D'Orazio and Lin 2022) .

Combining these data with the data provided by VIEWS gives a total of 74 predictors. The different classes are shown in Table 1, with the full set of features in the appendix.

Transforming the Dependent Variable

To produce different sets of point predictions for each grid-month, different Box-Cox transformations of the dependent variable were experimented with. The formula is shown in Equation 1. Values of λ that were tested include -1, -0.5, 0, 0.5, and 1.

Table 1. Variable Classes

Class	Description
Violence	Thresholds, counts, measures of violence at time t.
Time since	Number of periods since violence has been observed.
Time since spatial	Number of periods since violence has been observed in a neighbor.
Time since decay	Function to weight temporal distance since violence.
Time lag	Violence at period $t - i$ where $i > 0$.
Spatial lag	Violence in neighbors, neighbors of neighbors, etc.
Space-time distance	Spatial and temporal distance to violence as a single number.
Time lag spatial lag	Violence in neighbors at period $t - i$ where $i > 0$.
Rolling max	Rolling maximum of violence in a time window.
Onset	Onset of violence in a time window.

$$f_\lambda(y) = \begin{cases} \dfrac{y^\lambda - 1}{\lambda} & for\ \lambda \neq 0 \\ log(y) & for\ \lambda = 0 \end{cases} \tag{1}$$

For each grid-month, the dependent variable was transformed, the model was trained, point predictions were obtained, and then the predictions were inverted back to the original count scale. The *dorazio_log* model corresponds to $\lambda = 0$. The *dorazio_ensemble* model uses different values of λ for different grid-months. This will be discussed in more detail further down.

Learning Algorithm

Following D'Orazio and Lin (2022), the H2O autoML system was used. Each run was given 2.5 hours of training time and a maximum number of 20 models. After completion, H2O also provided a stacked ensemble of all ten models, and a stacked ensemble of the best performing models from each family of algorithms. The performance metric for H2O was the root mean squared error. For each run, the best performing model was selected and used to generate point predictions for the grid-months it was trained to forecast for.

2 Forecast Distributions

Forecast distributions reflect uncertainty in the point estimates. The default distribution for the VIEWS competition is take 1,000 draws from the Poisson distribution where the point prediction is sole parameter, representing both the mean and variance. Any distribution can be used, and the ones experimented with here were the Negative Binomial and the Tweedie. There are two parameters for the Negative Binomial: the mean and the dispersion. There are three parameters for the Tweedie: the mean, dispersion, and the power.

The first approach at parameterizing these distributions was to draw 100 values for each grid-month over the last five years using the point prediction as the mean and a fixed value for the dispersion and power. These 100 draws were used to calculate the continuous rank probability score (CRPS). The CRPS was also calculated using the default Poisson approach. The Negative Binomial was consistently the best performing distribution.

The second approach attempted to improve on the first by allowing for grid-specific parameterizations. The results from these experiments showed that the Tweedie was consistently as good or better than the Negative Binomial. Thus, the Tweedie distribution was selected. The potential dispersion values were .1, .5, 1, 2, 3, 4, 5. The potential power values were 1, 1.25, 1.5, 1.75, 2. Thus, there were a total of 35 different possible parameterizations for each grid. Each of the 360 models (72 months * 5 Box-Cox transformations) has its own grid-specific parameterizations, from which 1,000 draws were taken.

3 Results

Table 2 shows the CRPS values by year for each of the Box-Cox transformations and the for the ensemble model. The *dorazio_log* model corresponds to $\lambda = 0$ in this table. While the CRPS is generally lower, the model tends to forecast relatively low values. However, bolder forecasts may be more valuable. The model where $\lambda = 1$ forecasts the highest values, but at a considerable cost to the CRPS.

Table 2. CRPS Values Across Models

λ	2018	2019	2020	2021	2022	2023	Mean
-1	0.1360	0.1109	0.1262	0.9335	1.1319	0.2187	0.4429
-0.5	0.1376	0.1094	0.1239	0.9316	1.1306	0.2178	0.4418
0	0.1439	0.1068	0.1221	0.9288	1.1304	0.2198	0.4420
0.5	0.1621	0.1062	0.1217	0.9297	1.1424	0.2770	0.4565
1	0.1635	0.1153	0.1290	0.9446	1.1598	0.7360	0.5414
Ensemble	0.1439	0.1151	0.1247	0.9318	1.1330	0.2332	0.4470

The *dorazio_ensemble* uses forecasts from the five different Box-Cox transformations, selecting which to use based on a separate grid-month forecasting model that was trained to predict when to use which transformation. The model does forecast higher values. For example, consider 2022. The forecast distribution for each grid-month consists of 1,000 draws. For *dorazio_log*, there are 39 grids with at least one draw greater than or equal to 100. For *dorazio_ensemble* there are 93. However, the higher forecasts come at a cost to the CRPS.

4 Appendix

	Predictors	Predictors
1	priogrid_gid	greq_2_sb_count
2	ged_sb	greq_3_sb_count
3	decay_ged_sb_1	greq_5_sb_count
4	decay_ged_sb_25	greq_10_sb_count
5	decay_ged_sb_5	ln_sb_count
6	decay_ged_sb_100	time_since_greq_1_sb_count
7	decay_ged_sb_500	time_since_greq_2_sb_count
8	splag_1_1_sb_1	time_since_greq_3_sb_count
9	splag_1_decay_ged_sb_1	time_since_greq_5_sb_count
10	treelag_1_sb	time_since_greq_10_sb_count
11	treelag_2_sb	tlag_1_sb_count
12	sptime_dist_k1_ged_sb	tlag_2_sb_count
13	sptime_dist_k10_ged_sb	tlag_3_sb_count
14	sptime_dist_k001_ged_sb	tlag_4_sb_count
15	ged_sb_splag_1	tlag_5_sb_count
16	mov_avg_6_ged_best_sb	tlag_6_sb_count
17	mov_avg_12_ged_best_sb	tlag_7_sb_count
18	mov_avg_36_ged_best_sb	tlag_8_sb_count
19	mov_sum_6_ged_best_sb	tlag_9_sb_count
20	mov_sum_12_ged_best_sb	tlag_10_sb_count
21	mov_sum_36_ged_best_sb	tlag_11_sb_count
22	ged_sb_tlag_1	tlag_12_sb_count
23	ged_sb_tlag_2	splag_1_1_sb_count
24	ged_sb_tlag_3	tlag_1_splag_1_1_sb_count
25	ged_sb_tlag_4	tlag_2_splag_1_1_sb_count
26	ged_sb_tlag_5	tlag_3_splag_1_1_sb_count
27	ged_sb_tlag_6	greq_1_splag_1_1_sb_count
28	ged_sb_tlag_7	greq_2_splag_1_1_sb_count
29	ged_sb_tlag_8	greq_3_splag_1_1_sb_count
30	ged_sb_tlag_9	greq_5_splag_1_1_sb_count
31	ged_sb_tlag_10	greq_10_splag_1_1_sb_count
32	ged_sb_tlag_11	time_since_greq_1_splag_1_1_sb_count
33	ged_sb_tlag_12	time_since_greq_2_splag_1_1_sb_count
34	ged_sb_decay_12_time_since	time_since_greq_3_splag_1_1_sb_count
35	ged_sb_tlag_1_splag_1	time_since_greq_5_splag_1_1_sb_count
36	sb_count	time_since_greq_10_splag_1_1_sb_count
37	greq_1_sb_count	gwno

References

D'Orazio, V. and Y. Lin. 2022. "Forecasting conflict in Africa with automated machine learning systems." *International Interactions 48*(4): 714–738.

Hegre, H., C. Bell, M. Colaresi, M. Croicu, F. Hoyles, R. Jansen, M. R. Leis, A. Lindqvist-McGowan, D. Randahl, E. G. Rød, et al. 2021. "ViEWS2020: Revising and evaluating the ViEWS political Violence Early-Warning System." *Journal of peace research 58*(3): 599–611.

Hegre, H., M. Croicu, K. Eck, and S. Högbladh. 2020. "Introducing the ucdp candidate events dataset." *Research & Politics 7*(3): 2053168020935257.

Hegre, H., P. Vesco, and M. Colaresi. 2022. "Lessons from an escalation prediction competition." *International Interactions 48*(4): 521–554.

Sundberg, R. and E. Melander. 2013. "Introducing the UCDP Georeferenced Event Dataset." *Journal of Peace Research 50*(4): 523–532.

Tollefsen, A. F., H. Strand, and H. Buhaug. 2012. "PRIO-GRID: A unified spatial data structure." *Journal of Peace Research 49*(2): 363–374.